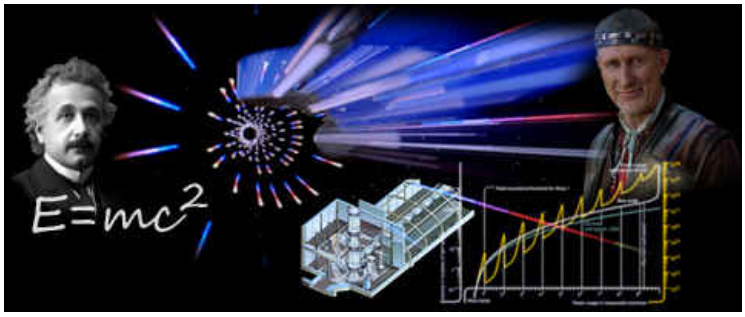
# The Physics and Technology of Warp Propulsion

by Bernd Schneider
http://www.ex-astris-scientia.org




## Introduction

***The Physics and Technology of Warp Propulsion***

*The Physics and Technology of Warp Propulsion*

- Introduction
- Chapter 1 - Real Physics and Interstellar Travel
- Chapter 3 - Subspace
- Chapter 6 - Warp Speed Measurement
- Chapter 7 - Appendix

"Nothing can move faster than light". This is a commonly known outcome of Albert Einstein's Special Theory of Relativity. Nevertheless, scientists as well as science fiction authors have been pondering about the possibility of faster-than-light (FTL) travel and its consequences ever since the theory became popular.

It is obvious that FTL propulsion is essential to allow interstellar travel in the first place. The nearest star system to Earth, for instance, is Proxima Centauri, roughly four light years away. Even if it were possible to achieve a slower-than-light (STL) speed very close to c (speed of light) in an extremely optimistic assumption, a one-way trip would take as long as four years. More realistic (and even much slower) STL travel times are given in the example of relativistic travel. Unless science fiction authors impose on themselves a restriction to cryogenic space travel or generational ships (in a subgenre often called "hard SF"), they need an explanation or justification why their characters can reach distant regions of space in a matter of days or hours.

Star Trek is certainly no "hard SF", still it ranks among the science fiction franchises with the most elaborated science and technology. Yet, the well-known warp drive is rather a visualization of FTL travel for TV than a real physical or even technological concept. And although Star Trek's warp drive frequently serves to illustrate hypothetical concepts of FTL propulsion in popular media, those hypotheses usually have nothing in common with the idea of enveloping a starship in a mass-lowering and propulsive warp field.

This article series is an attempt to summarize the most important facts about FTL travel in general and in real science, and about and warp drive as presented in Star Trek in particular. The Warp Propulsion series is less homogeneous than other parts of this website, as it has to take into account real-world physics as well as canon Star Trek facts, and carefully combine them without too much conjecture, to show the whole picture.

Fig. 0.1 Warp effect in "Star Trek: The Motion Picture"

Fig. 0.2 Warp effect in Star Trek: The Next Generation

Chapter 1 - Real Physics and Interstellar Travel is a strictly scientific account on real-world physics. It summarizes the underlying principles of movements at very fast speeds, in particular special relativity, which imposes limits on real-world propulsion systems. Hypothetical concepts to overcome these limitations are outlined in Chapter 2. Note that none of them could be developed to a viable theory so far, much less to a technical solution. Chapter 3 - Subspace is a treatise on subspace, the domain in which warp travel takes place in Star Trek. As opposed to chapter 1, there is not really a scientific basis for subspace; therefore the main scope of chapter 3 is the collection of canon evidence and its interpretation. The same applies to chapter 4 and 5 which deal with the history and the technical implementation of warp propulsion as depicted in Star Trek. Chapter 6 - Warp Speed Measurement analyzes the measurement of warp speeds and attempts to explain the different scales. Chapter 7 - Appendix, finally, provides a collection of constants, a glossary of terms and a reference list.

Fig. 0.3 Warp effect in "Star Trek: First Contact"

Fig. 0.4 Warp effect in "Star Trek Into Darkness"

## 1 Real Physics and Interstellar Travel

## 1.1 Classical Physics

*The Physics and Technology of Warp Propulsion*

- Introduction
- Chapter 1 - Real Physics and Interstellar Travel
- Chapter 3 - Subspace
- Chapter 6 - Warp Speed Measurement
- Chapter 7 - Appendix

This chapter summarizes some very basic theorems of physics. They mostly predate the theories of special relativity and of general relativity.

**Newton's laws of motion** Isaac Newton discovered the following laws. They are still valid (meaning that they are a very good approximation) for speeds much slower than the speed of light.

**1.** An object at rest or in uniform motion in a straight line will remain at rest or in the same uniform motion, respectively, unless acted upon by an unbalanced force. This is also known as the law of inertia.

**2.** The acceleration a of an object is directly proportional to the total unbalanced force F exerted on the object, and is inversely proportional to the mass m of the object (in other words, as mass increases, the acceleration has to decrease). The acceleration of an object has the same direction as the resulting force. This is also known as the law of acceleration.

$$F = m \cdot a$$

Eq. 1.1

**3.** If one object exerts a force on a second object, the second object exerts a force equal in magnitude and opposite in direction on the object body. This is also known as the law of interaction.

**Gravitation** Two objects with a mass of $m_1$ and $m_2$, respectively, and a distance of r between the centers of mass, attract each other with a force F of:

$$F = G\frac{m_1 m_2}{r^2}$$

Eq. 1.2

$G=6.67259*10^{-11}m^3kg^{-1}s^{-2}$ is Newton's constant of gravitation. If an object with a mass m of much less than Earth's mass is close to Earth's surface, it is convenient to approximate Eq. 1.2 as follows:

$$F = m \cdot g$$

Eq. 1.3

Here g is an acceleration slightly varying throughout Earth's surface, with an average of $9.81ms^{-2}$.

**Momentum conservation** In an isolated system, the total momentum is constant. This fundamental law is not affected by the theories of Relativity.

**Energy conservation** In an isolated system, the total energy is constant. This fundamental law is not affected by the theories of Relativity.

**Second law of thermodynamics** The overall entropy of an isolated system is always increasing. Entropy generally means disorder. An example is the heat flow from a warmer to a colder object. The entropy in the colder object will increase more than it will decrease in the warmer object. This why the reverse process, leading to lower entropy, would never take place spontaneously.

**Doppler shift** If the source of the wave is moving relative to the receiver or the other way round, the received signal will have a different frequency than the original signal. In the case of sound waves, two cases have to be distinguished. In the first case, the signal source is moving with a speed v relative to the medium, mostly air, in which the sound is propagating at a speed w:

$$f = f_0 \frac{1}{1 \pm \frac{v}{w}}$$

Eq. 1.4

f is the resulting frequency, $f_0$ the original frequency. The plus sign yields a frequency decrease in case the source is moving away, the minus sign an increase if the source is approaching. If the receiver is moving relative to the air, the equations are different. If v is the speed of the receiver, then the following applies to the frequency:

$$f = f_0\left[1 \pm v / w\right]$$

Eq. 1.5

Here the plus sign denotes the case of an approaching receiver and an according frequency increase; the minus sign applies to a receiver that moves away, resulting in a lower frequency.

The substantial difference between the two cases of moving transmitter and moving receiver is due to the fact that sound needs air in order to propagate. Special relativity will show that the situation is different for light. There is no medium, no "ether" in which light propagates and the two equations will merge to one relativistic Doppler shift.

**Particle-wave dualism** In addition to Einstein's equivalence of mass and energy (see 1.2 Special Relativity), de Broglie unified the two terms in that any particle exists not only alternatively but even simultaneously as matter and radiation. A particle with a mass m and a speed v was found to be equivalent to a wave with a wavelength lambda. With h, Planck's constant, the relation is as follows:

$$\lambda = \frac{h}{mv}$$

Eq. 1.6

The best-known example is the photon, a particle that represents electromagnetic radiation. The other way round, electrons, formerly known to have particle properties only, were found to show a diffraction pattern which would be only possible for a wave. The particle-wave dualism is an important prerequisite to quantum mechanics.

## 1.2 Special Relativity

Special relativity (SR) doesn't play a role in our daily life. Its impact becomes apparent only for relative speeds that are considerable fractions of the speed of light, c. I will henceforth occasionally refer to them as "relativistic speeds". The effects were first measured as late as towards the end of the 19th century and explained by Albert Einstein in 1905.

> **Side note** There are many approaches in literature and in the web to explain special relativity. Please refer to the appendix. A very good reference from which I have taken several suggestions is Jason Hinson's article on Relativity and FTL travel [Hin1]. You may wish to read his article in parallel.

The whole theory is based on two postulates:

> **1.** There is no invariant "fabric of space" or "ether" relative to which an absolute speed could be defined or measured. The terms "moving" or "resting" make only sense if they refer to a certain other frame of reference. The perception of movement is always mutual; the starship pilot who leaves Earth could claim that he is actually resting while the solar system is moving away.
>
> **2.** The speed of light, $c=3*10^8$m/s in the vacuum, is the same in all directions and in all frames of reference. This means that nothing is added or subtracted to this speed as the light source apparently moves.

**Frames of reference** In order to explain special relativity, it is crucial to introduce frames of reference. Such a frame of reference is basically a point-of-view, something inherent to an individual observer who perceives an event from a certain angle. The concept is in some way similar to the trivial spatial parallax where two or more persons see the same scene from different spatial angles and therefore give different descriptions of it. However, the following considerations are somewhat more abstract. "Seeing" or "observing" will not necessarily mean a sensory perception. On the contrary, the observer is assumed to take into consideration every "classic" measurement error such as signal delay or Doppler shift.

Aside from these effects that can be rather easily handled mathematically there is actually one even more severe constraint. The considerations on special relativity require inertial frames of reference. According to the definition in the general relativity chapter, this would be a floating or free falling frame of reference. Any presence of gravitational or acceleration forces would not only spoil the measurement, but could ultimately even question the validity of the SR. One provision for the following considerations is that all observers should float within their starships in space so that they can be regarded as local inertial frames. Basically, every observer has their own frame of reference; two observers are in the same frame if their relative motion to a third frame is the same, regardless of their distance.

**Space-time diagram** Special relativity takes place in the four-dimensional space-time. In an inertial frame all the Cartesian spatial coordinates x, y and z are equivalent (there is no "up" and "down" in space in a total absence of gravitational forces). Hence, for the sake of simplicity, we may substitute the three axes with one generic horizontal space (x-) axis which may stand for any of the coordinates x,y,z or all of them. Together with the vertical time (t-) axis we obtain a two-dimensional diagram (Fig. 1.1). It is very convenient to give the distance in light-years and the time in years. Irrespective of the current frame of reference, the speed of light always equals c and would be exactly 1ly per year in our diagram, according to the second postulate. Any light beam will therefore always form an angle of either 45° or -45° with the x-axis and the t-axis, as indicated by the yellow lines.

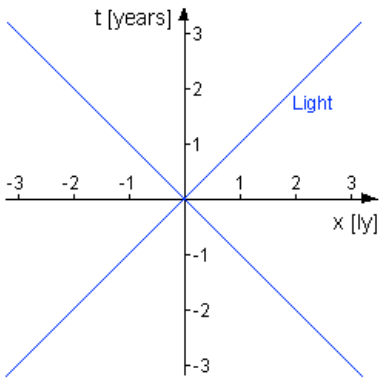



Fig. 1.1 Space-time diagram for a resting observer

A resting observer O (or better: an observer who thinks to be resting) draws an exactly perpendicular x-t-diagram for himself. The x-axis is equivalent to t=0 and is therefore a line of simultaneity, meaning that for O everything located on this line is simultaneous. This applies to every horizontal line t=const. likewise. The t-axis and every line parallel to it denote x=const. and therefore no movement in this frame of reference. If O is to describe the movement of another observer O* with respect to himself, O*'s time axis t* is sloped, and the reciprocal slope indicates a certain speed v=x/t. Fig. 1.2 shows O's coordinate system in gray, and O*'s in O's view in white. The common origin denotes that O and O* pass by each other (x=x*=0) at the time t=t*=0. At the first glance it seems strange that O*'s x*-axis is sloped into the opposite direction than his t*-axis, distorting his coordinate system.

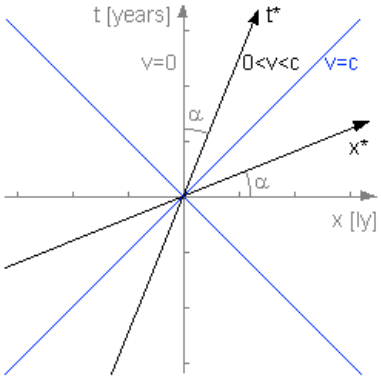



Fig. 1.2 Space-time diagram for a resting and a moving observer

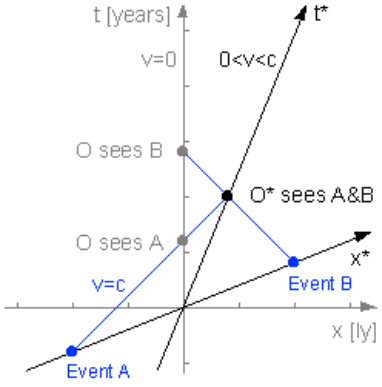



Fig. 1.3 Sloped x*-axis of a moving observer

The x*-axis can be explained by assuming two events A and B occurring at t*=0 at some distance to the two observers, as depicted in Fig. 1.3. O* sees them simultaneously, whereas O sees them at two different times. Since the two events are simultaneous in O*'s frame, the line A-0-B defines his x*-axis. A and B might be located anywhere on the 45-degree light paths traced back from the point "O* sees A&B", so we need further information to actually locate A and B. Since O* is supposed to actually *see* them at the same time (and not only date them back to t*=0), we also know that the two events A and B must have the same distance from the origin of the coordinate system. Now A and B are definite, and connecting them gives us the x*-axis. Some simple trigonometry would reveal that actually the angle between x* and x is the same as between t* and t, only the direction is opposite.

The faster the moving observer is, the closer will the two axes t* and x* move to each other. It becomes obvious in Fig. 1.2 that finally, at v=c, they will merge to one single axis, equivalent to the path of a light beam.

**Time dilation** The above space-time diagrams don't have a scale on the x*- and t*-axes so far. It is not the same scale as on the x- and t-axis, respectively. The method of determining the t*-scale is illustrated in the upper left half of Fig. 1.4. When the moving observer O* passes the resting observer O, they both set their clocks to t*=0 and t=0, respectively. Some time later, O's clock shows t=3, and at a yet unknown instant O*'s clock shows t*=3. The yellow light paths show when O will actually *see* O*'s clock at t*=3 (which is obviously after his own t=3 because the light needs time to travel), and vice versa. If O is smart enough, he can calculate the time when this light was emitted (by tracing back the -45° yellow line to O*'s t*-axis). His lines of simultaneity are exactly horizontal (red line), and the reconstructed event "t*=3" will take place at some yet unknown time on his t-axis. And even though O has taken into account the speed of light, "t*=3" still happens later than t=3. The quotation marks distinguish O's reconstruction of "t*=3" and O*'s direct reading t*=3. O* will do the same by reconstructing the event "t=3" (green line), which happens after his own t*=3. In other words, just as O sees the time of O* dilated, O* can say the same about O's time.

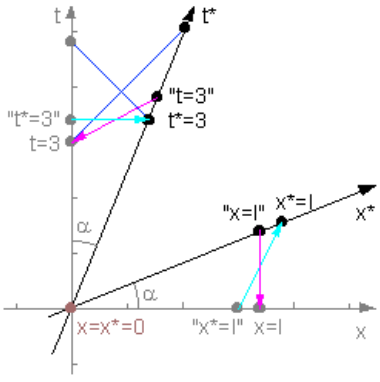



Fig. 1.4 Illustration of time dilation (upper left half) and length contraction (lower right half)

Since O*'s x*-axis and therefore his green line of simultaneity is sloped while the x-axis is not, it is impossible that any two events t=3 and "t*=3" and the two corresponding events t*=3 and "t=3" are simultaneous (and would occupy a single point) on the respective t-axis. If there is no absolute simultaneity, at least one of the two observers would see the other one's time dilated (in slow motion). Now we have to apply the first postulate, the principle of relativity. There must not be any preferred frame of reference, all observations have to be mutual. Either observer sees the other one's time dilated by the same factor. Not only in this example, but always. In our diagram the red and the green line have to cross to fulfill this postulate. More precisely, the ratio "t*=3" to t=3 has to be equal to "t=3" to t*=3. Some further calculations yield the following time dilation:

$$\frac{t}{t^*} = \sqrt{1-\tan^2\alpha} = \sqrt{1-\frac{v^2}{c^2}} = \frac{1}{\text{gamma}}$$

Eq. 1.7

**Side note** When drawing the axes to scale in an x-t diagram, one has to account for the inherently longer hypotenuse t* and multiply the above formula with an additional factor cos alpha to "project" t* on t.

Note that the time dilation would be the square root of a negative number (imaginary) if we assume an FTL speed v>c. Imaginary numbers are not really forbidden, on the contrary, they play an important role in the description of waves. Anyway, a physical quantity such as the time doesn't make any sense once it gets imaginary. Unless a suited interpretation or a more comprehensive theory is found, considerations end as soon as a time (dilation), which has to be finite and real by definition, would become infinitely large or imaginary. The same applies to the length contraction and mass increase. Star Trek's warp theory (although it does not really have a mathematical description) circumvents all these problems in a way that no such relativistic effects occur.

**Length contraction** The considerations for the scale of the x*-axis are similar as those for the t*-axis. They are illustrated in the bottom right portion of Fig. 1.4. Let us assume that O and O* both have identical rulers and hold their left ends x=x*=0 when they meet at t=t*=0. Their right ends are at x=l on the x-axis and at x*=l on the x*-axis, respectively. O and his ruler rest in their frame of reference. At t*=0 (which is not simultaneous with t=0 at the right end of the ruler!) O* obtains a still unknown length "x=l" for O's ruler (green line). O* and his ruler move along the t*-axis. At t=0, O sees an apparent length "x=l" of O*'s ruler (red line). Due to the slope of the t*-axis, it is impossible that the two observers mutually see the same length l for the other ruler. Since the relativity principle would be violated in case one observer saw two equal lengths and the other one two different lengths, the mutual length contraction must be the same. Note that the geometry is virtually the same as for the time dilation, so it's not astounding that length contraction is determined by the factor gamma too:

$$\frac{x^*}{x} = \sqrt{1-\tan^2\alpha} = \sqrt{1-\frac{v^2}{c^2}} = \frac{1}{\text{gamma}}$$

Eq. 1.8

**Side note** Once again, note that when drawing the x*-axis to scale, a correction is necessary, a factor of cos alpha to the above formula.

**Addition of velocities** One of the most popular examples used to illustrate the effects of special relativity is the addition of velocities. It is obvious that in the realm of very slow speeds it's possible to simply add or subtract velocity vectors from each other. For simplicity, let's assume movements that take place in only one dimension so that the vector is reduced to a plus or minus sign along with the absolute speed figure, like in the space-time diagrams. Imagine a tank that has a speed of v compared to the ground and to an observer standing on the ground (Fig. 1.5). The tank fires a projectile whose speed is determined as w by the tank driver. The resting observer, on the other hand, will register a projectile speed of v+w relative to the ground. So far, so good.

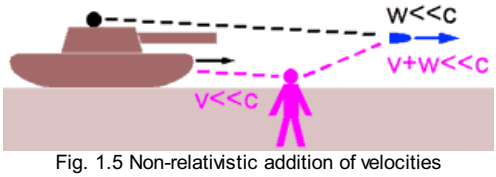



Fig. 1.5 Non-relativistic addition of velocities

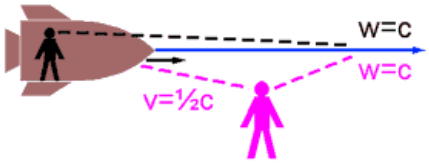



Fig. 1.6 Relativistic addition of velocities

The simple addition (or subtraction, if the speeds have opposite directions) seems very obvious, but it isn't so if the single speeds are considerable fractions of c. Let's replace the tank with a starship (which is intentionally a generic vessel, no Trek ship), the projectile with a laser beam and assume that both observers are floating, one in open space and one in his uniformly moving rocket (no acceleration), at a speed of c/2 compared to the first observer (Fig. 1.6). The rocket pilot will see the laser beam move away at exactly c. This is still exactly what we expect. However, the observer in open space won't see the light beam travel at v+c=1.5c but only at c. Actually, any observer with any velocity (or in any frame of reference) would measure a light speed of exactly $c=3*10^8$m/s.

Space-time-diagrams allow to derive the addition theorem for relativistic velocities. The resulting speed u is given by:

$$u = \frac{v+w}{1+\frac{vw}{c^2}}$$

Eq. 1.9

For v,w<<c we may neglect the second term in the denominator. We obtain u=v+w as we expect it for small speeds. If vw gets close to $c^2$, the speed u may be close to, but never equal to or even faster than c. Finally, if either v or w equals c, u is equal to c as well. There is obviously something special to the speed of light. c always remains constant, no matter where in which frame and which direction it is measured. c is also the absolute upper limit of all velocity additions and can't be exceeded in any frame of reference.

**Mass increase** Mass is a property inherent to any kind of matter. We may distinguish two forms of mass, one that determines the force that has to be applied to accelerate an object (inert mass) and one that determines which force it experiences in a gravitational field (heavy mass). At latest since the equivalence principle of GR they have been found to be absolutely identical.

However, mass is apparently not an invariant property. Consider two identical rockets that lifted off together at t=0 and now move away from the launch platform in opposite directions, each with a constant absolute speed of v. Each pilot sees the launch platform move away at v, while Eq. 1.9 shows us that the two ships move away from each other at a speed u<2v. The "real" center of mass of the whole system of the two ships would be still at the launch platform, however, each pilot would see a center of mass closer to the other ship than to his own. This may be interpreted as a mass increase of the other ship to m compared to the rest mass $m_0$ measured for both ships prior to the launch:

$$m = \frac{m_0}{\sqrt{1-\frac{v^2}{c^2}}} = m_0 \cdot \text{gamma}$$

Eq. 1.10

This function is plotted in Fig. 1.7.

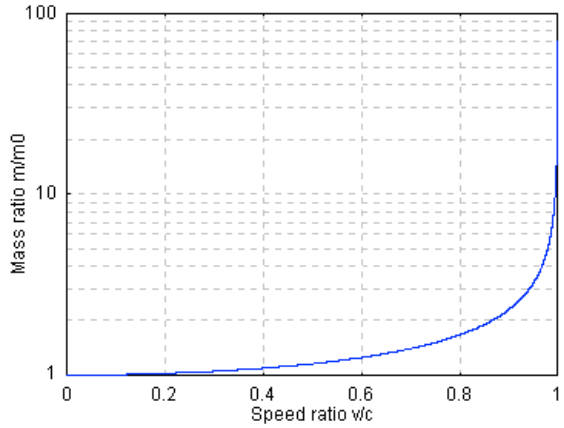



Fig. 1.7 Mass increase for relativistic speeds

So each object has a rest mass $m_0$ and an additional mass $m-m_0$ due to its speed as seen from another frame of reference. This is actually a convenient explanation for the fact that the speed of light cannot be reached. The mass increases more and more as the object approaches c, and so would the required momentum to propel the ship.

Finally, at v=c, we would get an infinite mass, unless the rest mass $m_0$ is zero. The latter must be the case for photons which always move at the speed of light, which even define the speed of light. If we assume an FTL speed v>c, the denominator will be the square root of a negative number, and therefore the whole mass will be imaginary. As already stated for the time dilation, there is not yet a suitable theory how an imaginary mass could be interpreted.

**Mass-energy equivalence** Let us consider Eq. 1.10 again. It is possible to express it as follows:

$$m = \frac{m_0}{\sqrt{1-\frac{v^2}{c^2}}} = m_0 + \frac{1}{2}m_0\frac{v^2}{c^2} + \frac{3}{8}m_0\frac{v^4}{c^4} + \ldots$$

Eq. 1.11

It is obvious that we may neglect the third order and the following terms for slow speeds. If we multiply the equation with $c^2$ we obtain the familiar Newtonian kinetic energy $\frac{1}{2}m_0v^2$ plus a new term $m_0c^2$. Obviously already the resting object has a certain energy content $m_0c^2$. We get a more general expression for the complete energy E contained in an object with a rest mass $m_0$ and a moving mass m, so we may write (drumrolls!):

$$E = mc^2$$

Eq. 1.12

Energy E and mass m are equivalent; the only difference between them is the constant factor $c^2$. If there is an according energy to each given mass, can the mass be converted to energy? The answer is yes, and Trek fans know that the solution is a matter/antimatter reaction in which the two forms of matter annihilate each other, thereby transforming their whole mass into energy.

**Light cone** Let us have a look at Fig. 1.1 again. There are two light beams running through the origin of the diagram, one traveling in positive and one in negative x direction. The slope is 1ly per year, which equals c. If nothing can move faster than light, then every t*-axis of a moving observer and every path of a message sent from one point to another in the diagram must be steeper than these two lines. This defines an area "inside" the two light beams for possible signal paths originating at or going to (x=0,t=0). This area is marked dark green in Fig. 1.8. The black area is "outside" the light cone. The origin of the diagram marks "here (x=0)" and "now (t=0)" for the resting observer.

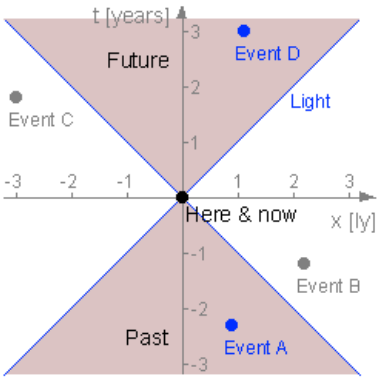



Fig. 1.8 The light cone

The common-sense definition tells us that "future" is any event at t>0, and past is any event at t<0. Special relativity shows us a different view of these two terms. Let us consider the four marked events which could be star explosions (novae), for instance. Event A is below the x-axis and within the light cone. It is possible for the resting observer O to see or to learn about the event in the past, since a -45° light beam would reach the t-axis at about one and a half years prior to t=0. Therefore this event belongs to O's past. Event B is also below the x-axis, but outside the light cone. The event has no effect on O in the present, since the light would need almost another year to reach him. Strictly speaking, B is not in O's past. Similar considerations are possible for the term "future". Since his signal wouldn't be able to reach the event C, outside the light cone, in time, O is not able to influence it. It's not in his future. Event D, on the other hand, is inside the light cone and may therefore be caused or influenced by the observer.

What about a moving observer? One important consequence of the considerations in this whole chapter was that two different observers will disagree about where and when a certain event happens. The light cone, on the other hand, remains the same, irrespective of the frame of reference. So even if two observers meeting at t=0 have different impressions about simultaneity, they will agree that there are certain, either affected (future) or affecting (past) events inside the light cone, and outside events they shouldn't bother about.

## 1.3 Twin Paradox

The considerations about the time dilation in special relativity had the result that the terms "moving observer" and "resting observer" are interchangeable as are their space-time diagrams. If there are two observers with a speed relative to each other, either of them will see the other one move. Either observer will see the other one's clock ticking slower. Special relativity necessarily requires that the observations are mutual, since it forbids a preferred, absolutely resting frame of reference. Either clock is slower than the other one? How is this possible?

**The problem** Specifically, the twin paradox is about twins of whom one travels to the stars at a relativistic speed, while the other one stays on Earth. It is obvious that the example assumes twins, since it would be easier to see if one of them actually looks older than the other one when they meet again. Anyway, it should work with unrelated persons as well. What happens when the space traveler returns to Earth? Is he the younger one, or maybe his twin on Earth, or are they equally old?

> **Side note** The following example for the twin paradox deliberately uses the same figures as Jason Hinson's excellent treatise on Relativity and FTL travel [Hin1], to increase the chance of understanding it. And because I was too lazy to recalculate everything. ;-)

To anticipate the result, the space traveler will be the younger one when he returns. The solution is almost trivial. Time dilation only remains the same as long as both observers stay in their respective frames of reference. However, if the two observers want to meet again, one of them or both of them have to change their frame(s) of reference. In this case it is the space traveler who has to decelerate, turn around, and accelerate his starship in the other direction. It is important to note that the whole effect can be explained without referring to any general relativity effects. Time dilation attributed to acceleration or gravity will change the result, but it will not play a role in the following discussion. The twin paradox is not really paradox, it can be solved, and this is best done with a space-time diagram.

**Part 1: moving away from Earth** Fig. 1.9 shows the first part of the FTL travel. O is the "resting" observer who stays on Earth the whole time. Earth is subsequently regarded as an approximated inertial frame. Strictly speaking, O would have to float in Earth's orbit, according to the definition in general relativity. Once again, however, it is important to say that the following considerations don't need general relativity at all. I only refer to O as staying in an inertial frame so as to exclude any GR influence.

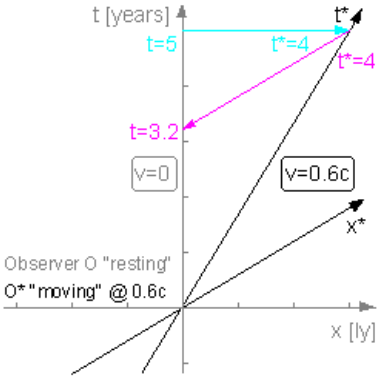


Fig. 1.9 Illustration of the twin paradox, moving away from Earth

The moving observer O* is supposed to travel at a speed of 0.6c relative to Earth and O. When O* passes by O (x=x*=0), they both set their clocks to zero (t=t*=0). So the origin of their space-time diagrams is the same, and the time dilation will become apparent in the different times t* and t for simultaneous events. As outlined above, t* is sloped, as is x* (see also Fig. 1.3). The measurement of time dilation works as outlined in Fig. 1.4. O's lines of simultaneity are parallel to his x-axis and perpendicular to his t-axis. He will see that 5 years on his t-axis correspond with only 4 years on the t*-axis (red arrow), because the latter is stretched according to Eq. 1.7. Therefore O*'s clock is ticking slower from O's point-of-view. The other way round, O* draws lines of simultaneity parallel to his sloped x*-axis and he reckons that O's clock is running slower, 4 years on his t*-axis compared to 3.2 years on the t-axis (green arrow). It is easy to see that the mutual dilation is the same, since 5/4 equals 4/3.2. Who is correct? Answer: Both of them, since they are in different frames of reference, and they stay in these frames. The two observers just see things differently; they wouldn't have to care whether their perception is "correct" and the other one is actually aging slower -- unless they wanted to meet again.

**Part 2: resting in space** Now let us assume that O* stops his starship when his clock shows t*=4 years, maybe to examine a phenomenon or to land on a planet. According to Fig. 1.10 he is now resting in space relative to Earth and his new x**-t** coordinate system is parallel to the x-t system of O on Earth. O* is now in the same frame of reference as O. And this is exactly the point: O*'s clock still shows 4 years, and he notices that not 3.2 years have elapsed on Earth as briefly before his stop, but 5 years, and this is exactly what O says too. Two observers in the same frame agree about their clock readings. O* has been in a different frame of reference at 0.6c for 4 years of his time and 5 years of O's time. This difference becomes a permanent offset when O* enters O's frame of reference. Paradox solved.

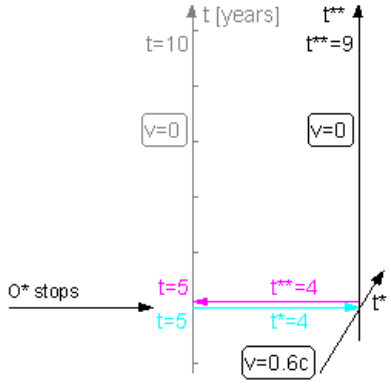



Fig. 1.10 Illustration of the twin paradox, landing or resting in space

It is obvious that the accumulative dilation effect will become the larger the longer the travel duration is. Note that O's clock has always been ticking slower in O*'s moving frame of reference. The fact that O's clock nevertheless suddenly shows a much later time (namely 5 instead of 3.2 years) is solely attributed to the fact that O* is entering a frame of reference in which exactly these 5 years have elapsed.

Once again, it is crucial to annotate that the process of decelerating would only change the result qualitatively, since there could be no exact kink, as O* changes from t* to t**. Practical deceleration is no sudden process, and the transition from t* to t** should be curved. Moreover, the deceleration itself would be connected with a time dilation according to GR, but the paradox is already solved without taking this into account.

**Part 3: return to Earth** Let us assume that at t*=4 years, O* suddenly gets homesick and turns around immediately instead of just resting in space. His relative speed is v=-0.6c during his travel back to Earth, the minus sign indicating that he is heading to the negative x direction. It is obvious that this second part of his travel should be symmetrical to the first part at +0.6c in Fig. 1.9, the symmetry axis being the clock comparison at t=5 years and t**=4 years. This is exactly the moment when O* has covered both half of his way and half of his time.

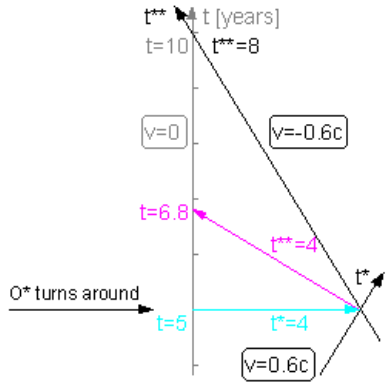



Fig. 1.11 Illustration of the twin paradox, return to Earth

Fig. 1.11 demonstrates what happens to O*'s clock comparison. Since he is changing his frame of reference from v=0.6c to v=-0.6c relative to Earth, the speed change and therefore the effect is twice as large as in Fig. 1.10. Assuming that O* doesn't stop for a clock comparison as he did above, he would see that O's clock directly jumps from 3.2 years to 6.8 years. Following O*'s travel back to Earth, we see that the end time is t**=8 years (O*'s clock) and t=10 years (O's clock). The traveling twin is actually two years younger at their family reunion.

We could imagine several other scenarios in which O might catch up with the traveling O*, so that O is actually the younger one.

Alternatively, O* could stop in space, and O could do the same travel as O*, so that they would be equally old when O reaches O*. The analysis of the twin paradox shows that the simple statement "moving observers age slower" is insufficient. The statement has to be modified in that "moving observers age slower as seen from any different frame of reference, and they notice it when they enter this frame themselves".

## 1.4 Causality Paradox

As already stated further above, two observers in different frames of reference will disagree about the simultaneity of certain events (see Fig. 1.3). What if the same event were in one observer's future, but in another observer's past when they meet each other? This is not a problem in special relativity, since no signal is allowed to travel faster than light. Any event that could be theoretically influenced by one observer, but has already happened for the other one, is outside the light cone depicted in Fig. 1.8. Causality is preserved.

Fig. 1.12 depicts the space-time diagrams of two observers with a speed relative to each other. Let us assume the usual case that the moving observer O* passes by the resting observer O at t=t*=0. They agree about the simultaneity of this passing event, but not about any other event at t<>0 or t*<>0. Event A is below the t-axis, but above the t*-axis. This doesn't matter as long as the two can send and receive only STL signals. Event A is outside the light cone, and the argument would be as follows: A is in O*'s future (future in a wider sense), but he has no means of influencing it at t*=0, since his signal couldn't reach it in time. A is in O's past (past in a wider sense), but it doesn't matter because he can't know of it at t=0.

What would be different if either FTL travel or FTL signal transfer were possible? In this case we would be allowed to draw signal paths of less than 45 degrees steepness in the space-time diagram. Let us assume that O* is able to send an FTL signal to influence or to cause event A in the first place, just when the two observers pass each other. Note that this signal would travel at v>c in any frame of reference, and that it would travel back in time in O's frame, since it runs into negative t-direction in O's orthogonal x-t coordinate system, to an event that is in O's past. If O* can send an FTL signal to cause the event A, then a second FTL signal can be sent to O to inform him of A as soon as it has just happened. This signal would run at v>c in positive t-direction for O, but in negative t*-direction for O*. So the situation is exactly inverse to the first FTL signal. Now O is able to receive a message from O*'s future.

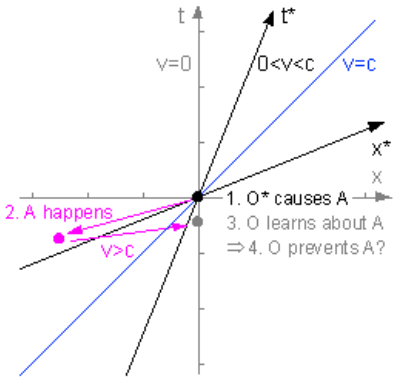



Fig. 1.12 Illustration of a possible causality paradox

The paradox occurs when O, knowing about the future, decides to prevent A from happening. Maybe O* is a bad guy, and event A is the death of an unfortunate victim, killed because of his FTL message. O would have enough time to hinder O*, to warn the victim or to take other precautions, since it is still t<0 when he receives the message, and O* has not yet caused event A.

The sequence of events (in logical rather than chronological order) would be as follows:

1. At t=t*=0, the two observers pass each other and O* sends an FTL message that causes A.
2. A happens in O*'s past (t*<0) and in O's future (t>0).
3. O learns about event A through another FTL signal, still at t<0, before he meets O*.
4. O might be able to prevent A from happening. However, how could O have learned about A if it actually never happened?

This is obviously the SR version of the well-known grandfather paradox. Note that these considerations don't take into account which method of FTL travel or FTL signal transfer is used. Within the realm of special relativity, they should apply to any form of FTL travel. Anyway, if FTL travel is feasible, then it is much like time travel. It is not clear how this paradox can be resolved. The basic suggestions are the same as for generic time travel and are outlined in my time travel article.

## 1.5 Other Obstacles to Interstellar Travel

**Power considerations** Rocket propulsion (as a generic term for any drive using accelerated particles) can be described by momentum conservation, resulting in the following simple equation:

$$m \cdot dv = -w \cdot dm$$

Eq. 1.13

The left side represents the infinitesimal speed increase (acceleration) dv of the ship with a mass m, the right side is the mass

decrease -dm of the ship if particles are thrusted out at a speed w. This would result in a constant thrust and therefore in a constant acceleration, at least in the range of ship speeds much smaller than c. Eq. 1.13 can be integrated to show the relation between an initial mass $m_0$, a final mass $m_1$ and a speed $v_1$ to be achieved:

$$-\frac{1}{w}\int_0^{v_1} dv = \int_{m_0}^{m_1} \frac{1}{m} dm \quad \Rightarrow \quad v_1 = w \cdot \ln\frac{m_0}{m_1} \quad \Rightarrow \quad m_1 = m_0 \cdot e^{-v_1/w}$$

Eq. 1.14

The remaining mass $m_1$ at the end of the flight, the payload, is only a fraction of the total mass $m_0$, the rest is the necessary fuel. The achievable speed $v_1$ is limited by the speed w of the accelerated particles, i.e. the principle of the drive, and by the fuel-to-payload ratio.

Let us assume a photon drive as the most advanced conventional propulsion technology, so that w would be equal to c, the speed of light. The fuel would be matter and antimatter in the ideal case, yielding an efficiency near 100%, meaning that according to Eq. 1.14 almost the complete mass of the fuel could contribute to propulsion. Eq. 1.13 and Eq. 1.14 would remain valid, with w=c. If relativistic effects are not yet taken into account, the payload could be as much as 60% of the total mass of the starship, if it's going to be accelerated to 0.5c. However, the mass increase at high sublight speeds as given in Eq. 1.10 spoils the efficiency of any available propulsion system as soon as the speed gets close to c, since the same thrust will effect a smaller acceleration. STL examples will be discussed in Section 1.7.

**Acceleration and Deceleration** Eq. 1.13 shows that the achievable speed is limited by the momentum (speed and mass) of the accelerated particles, provided that a conventional rocket engine is used. The requirements of such a drive, e.g. the photon drive outlined above, are that a considerable amount of particles has to be accelerated to a high speed at a satisfactory efficiency.

Even more restrictively, the human body simply couldn't sustain accelerations of much more than $g=9.81 ms^{-2}$, which is the acceleration on Earth's surface. Accelerations of several g are taken into account in aeronautics and astronautics only for short terms, with critical peak values of up to 20g. Unless something like Star Trek's IDF (inertial damping field) will be invented [Ste91], it is probably the most realistic approach to assume a constant acceleration of g from the traveler's viewpoint during the whole journey. This would have the convenient side effect that an artificial gravity equal to Earth's surface would be automatically created.

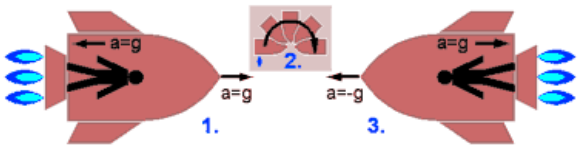



Fig. 1.13 Concept of the turn-around starship

According to Newton's first and second postulates it will be necessary to decelerate the starship as it approaches the destination. Thus, the starship needs a "brake". It wouldn't be very wise to install a second, equally powerful engine at the front of the starship for this purpose. Moreover, the artificial gravity would act in the opposite direction during the deceleration phase in this case. The alternative solution is simple: Half-way to the destination, the starship would be simply turned around by means of maneuvering thrusters so that it now decelerates at a rate of g; the artificial gravity would remain exactly the same. Actually, complying with the equivalence principle of general relativity, if the travelers didn't look out of the windows or at their sensor readings, they wouldn't even notice that the ship is now decelerating. Only during the turn-around the gravity would change for a brief time while the main engines are switched off. Fig. 1.13 depicts such a turn-around ship, "1." is the acceleration phase, "2." the turn-around, "3." the deceleration.

**Doppler shift** In the chapter on classical physics, the Doppler effect has been described as the frequency increase or decrease of a sound wave. The two cases of a moving source and moving observer only have to be distinguished in case of an acoustic signal, because the speed of sound is constant relative to the air and therefore the observer would measure different signal speeds in the two cases. Since the speed of light is constant, there is only one formula for Doppler shift of electromagnetic radiation, already taking into account SR time dilation:

$$f^* = f \cdot \sqrt{\frac{1 - v/c}{1 + v/c}}$$

Eq. 1.15

Note that Eq. 1.15 covers both cases of frequency increase (v and c in opposite directions, v/c<0) and frequency decrease (v and c in the same direction, v/c>0). Since the power of the radiation is proportional to its frequency, the forward end will be subject to a higher radiation power and dose than the rear end, assuming isotropic (homogeneous) radiation.

Actually, for STL travel the Doppler shift is not exactly a problem. At 0.8c, for instance, the radiation power at the bow is three times the average of the isotropic radiation. Visible light would "mutate" to UV radiation, but its intensity would still be far from dangerous. Only if v gets very close to c (or -c, to be precise), the situation could get critical for the space travelers, and an additional shielding would be necessary. On the other hand, it's useless anyway to get as close to c as possible because of the mass increase. For v=-c, the Doppler shift would be theoretically infinite.

It is not completely clear (but may become clear in one of the following chapters) how Doppler shift can be described for an FTL drive in general or warp propulsion in particular. It could be the conventional, non-relativistic Doppler shift that applies to warp drive since mass increase and time dilation are not valid either. In this case the radiation frequency would simply increase to 1+|v/c| times the original frequency, and this could be a considerable problem for high warp speeds, and it would require thick radiation shields and forbid forward windows.

## 1.6 General Relativity

As we can infer from the name, general relativity (GR) is a more comprehensive theory than special relativity. Although the concept as a whole has to be explained with massive use of complex mathematics, the basic principles are quite evident and perhaps easier to understand than those of special relativity. General relativity takes into account the influence of the presence of a mass, of gravitational fields caused by this mass.

**Inertial frames** The chapter on special relativity assumed inertial frames of reference, that is, frames of reference in which there is no acceleration to the object under investigation. The first thought might be that a person standing on Earth's surface should be in an inertial frame of reference, since he is not accelerated relative to Earth. This idea is wrong, according to GR. Earth's gravity "spoils" the possible inertial frame. Although this is not exactly what we understand as "acceleration", there can't be an inertial frame on Earth's surface. Actually, we have to extend our definition of an inertial frame.

**Principle of equivalence** Consider the rocket in the left half of Fig. 1.14 whose engines are powered somewhere in open space, far away from any star or planet. According to Newton's Second Law of Motion (see 1.1 Classical Physics), if the engine force is constant the acceleration will be constant too. The thrust may be adjusted in a way that the acceleration is exactly $g=9.81ms^{-2}$, equal to the acceleration in Earth's gravitational field. The passenger will then be able to stand on the rocket's bottom as if it were Earth's surface, since the floor of the rocket exerts exactly the same force on him in both cases. Compare it to the right half of Fig. 1.14; the two situations are equivalent. "Heavy mass" and "inert mass" are actually the same.

One might object that there should still be many differences. Specifically one should expect that a physicist who is locked up in such a starship (without windows) should be able to find out whether it is standing on Earth or accelerating in space. The surprising result of GR is that he will get exactly the same experimental results in both cases. Imagine that the rocket is quite long, and our physicist sends out a laser beam from the rocket's bottom to its top. In the case of the accelerating rocket we would not be surprised that the frequency of the light beam decreases, since the receiver would virtually move away from the source while the beam is on the way. This effect is the familiar Doppler shift. We wouldn't expect the light frequency (and therefore its intensity) to decrease inside a stationary rocket too, but that's exactly what happens in Earth's gravitational field! The light beam has to "climb up" in the field, thereby losing energy, which becomes apparent in a lower frequency. Obviously, as opposed to common belief so far, light is affected by gravity.

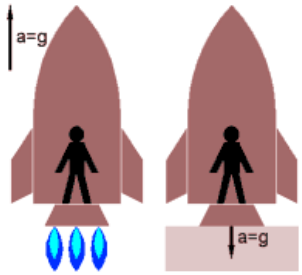



Fig. 1.14 Equivalence of rocket standing on Earth and rocket accelerated in space (no inertial frames)

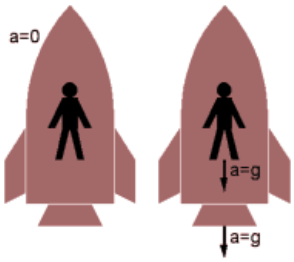



Fig. 1.15 Equivalence of free falling rocket and rocket floating in space (inertial frames)

Let us have a look at Fig. 1.15. The left half shows a rocket floating in space, far away from any star or planet. No force acts upon the passenger, he is weightless. It's not only a balance of forces, but all forces are actually zero. This is an inertial frame, or at least a very good approximation thereof. There can be obviously no perfect inertial frame as long as there is still a certain mass present. Compare this to the depiction of the free falling starship in the right half of Fig. 1.15. Both the rocket and the passenger are attracted with the same acceleration a=g. Although there is acceleration, this is an inertial frame too, and it is equivalent to the floating rocket. The point is that in both cases the inside of the rocket is an inertial frame, since the ship and passenger don't exert any force/acceleration on each other. About the same applies to a parabolic flight or a ship in orbit during which the passengers are weightless.

We might have found an inertial frame also in the presence of a mass, but we have to keep in mind that this can be only an approximation. Consider a very long ship falling down to Earth. The passenger in the rocket's top would experience a smaller acceleration than the rocket's bottom and would have the impression that the bottom is accelerated with respect to himself. Similarly, in a very wide rocket (it may be the same one, only turned by 90 degrees), two people at either end would see that the other one is accelerated towards him. This is because they would fall in slightly different radial directions to the center of mass. None of these observations would be allowed within an inertial frame. Therefore, we are only able to define local inertial frames.

**Time dilation** As already mentioned above, there is a time dilation in general relativity, because light will gain or lose potential energy when it is moving farther away from or closer to a center of mass, respectively. The time dilation depends on the gravitational potential as given in Eq. 1.2 and amounts to:

$$t^* = t \cdot \left[1 - \frac{2GM}{rc^2}\right]$$

Eq. 1.16

G is Newton's gravitational constant, M is the planet's mass, and r is the distance from the center of mass. Eq. 1.16 can be approximated in the direct vicinity of the planet using Eq. 1.3:

$$t^* \approx t \cdot \left[1 - \frac{2gh}{c^2}\right]$$

Eq. 1.17

In both equations t* is the time elapsing on the surface, while t is the time in a height h above the surface, with g being the standard acceleration. The time t* is always shorter than t so that, theoretically, people living on the sea level age slower than those in the mountains. The time dilation has been measured on normal plane flights and summed up to 52.8 nanoseconds when the clock on the plane was compared to the reference clock on the surface after 40 hours [Sex87]. 5.7 nanoseconds had to be subtracted from this result, since they were attributed to the time dilation of relativistic movements that was discussed in the chapter about special relativity.

**Curved space** The time dilation in the above paragraph goes along with a length contraction in the vicinity of a mass:

$$x^* = x \cdot \left[1 - \frac{2\,GM}{rc^2}\right]$$

Eq. 1.18

However, length contraction is not a commonly used concept in GR. The equivalent idea of a curved space is usually preferred. For it is obviously impossible to illustrate the distortions of a four-dimensional space-time, we have to restrict our considerations to two spatial dimensions. Imagine two-dimensional creatures living on an even plastic foil. The creatures might have developed a plane geometry that allows them to calculate distances and angles. Now someone from the three-dimensional world bends and stretches the plastic foil. The two-dimensional creatures will be very confused about it, since their whole knowledge of geometry doesn't seem to be correct anymore. They might apply correction factors in certain areas, compensating for points that are measured as being closer together or farther away from each other than their calculations indicate. Alternatively, a very smart two-dimensional scientist might come up with the idea that their area is actually not flat but warped. About this is what general relativity says about our four-dimensional space-time.

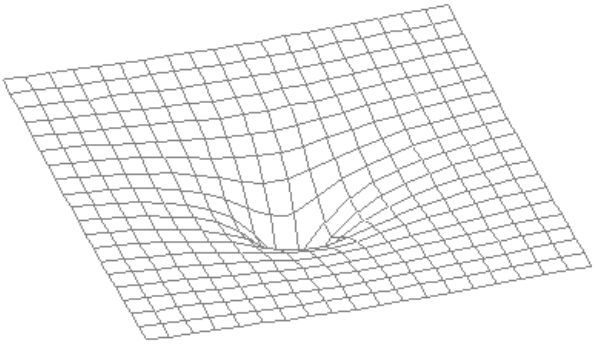


Fig. 1.16 Illustration of curved space

Fig. 1.16 is limited to the two spatial dimensions x and y. It can be regarded as something like a "cross-section" of the actual spatial distortion. We can imagine that the center of mass is somewhere in the middle "underneath" the x-y plane, where the curvature is most pronounced (a "gravity well").

**Speed of light** A light beam passing by an area with strong gravity such as a star will not be "straight", at least it will not appear straight as seen from flat space. Using an exact mathematical description of curved space, the light beam will follow a geodesic. Using a more illustrative idea, because of its mass the light beam will be deflected by a certain angle. The first reliable measurements were performed during a total solar eclipse in 1919 [Sex87]. They showed that the apparent positions of stars whose light was passing the darkened sun were farther away from the sun than the "real" positions measured in the night. It is possible to calculate the deflection angle assuming that light consists of particles and using the Newtonian theory of gravitation, however, this accounts for only half the measured value.

There is another effect involved that can only be explained with general relativity. As it is the case in materials with different refraction indices, light will "avoid" regions in which its apparent speed is reduced. A "detour" may therefore become a "shortcut". This is what happens in the vicinity of a star and what is responsible for the other 50% of the light deflection.

Now we can see the relation of time dilation, length contraction, the geometry of space and the speed of light. A light beam would have a definite speed c=r/t in the "flat" space in some distance from the center of mass. Closer to the center, space itself is "curved", and this again is equivalent to the effect that everything coming from outside would apparently "shrink". A ruler with a length r would appear shortened to r*. Since the time t* is shortened with respect to t by the same factor, c=r/t=r*/t* remains constant. This is what an observer inside the gravity well would measure, and what the external observer would confirm. On the other hand, the external observer would see that the light inside the gravity well takes a detour (judging from his geometry of flat space) *or* would pass a smaller effective distance in the same time (regarding the shortened ruler) *and* the light beam would be additionally slowed down because of the time dilation. Thus, he would measure that the light beam actually needs a longer time to pass by the gravity well than t=r/c. If he is sure about the distance r, then the effective c* inside the gravity well must be smaller:

$$c^* = c \cdot \left[1 - \frac{2\,GM}{rc^2}\right]^2$$

Eq. 1.19

It was confirmed experimentally that a radar signal between Earth and Venus takes a longer time than the distance between the planets indicates if it passes by close to the sun.

**Black holes** Let us have a look at Eq. 1.16 and Eq. 1.18 again. Obviously something strange happens at $r=2GM/c^2$. The time t* becomes zero (and the time dilation infinite), and the length x* is contracted to zero. This is the Schwarzschild radius or event horizon, a quantity that turns up in several other equations of GR. A collapsing star whose radius shrinks below this event horizon will become a

black hole. Specifically, the space-time inside the event horizon is curved in a way that every particle will inevitably fall into its center. It is unknown how dense the matter in the center of a black hole is actually compressed. The mere equations indicate that the laws of physics as we know them wouldn't be valid any more (singularity). On the other hand, it doesn't matter to the outside world what is going on inside the black hole, since it will never be possible to observe it.

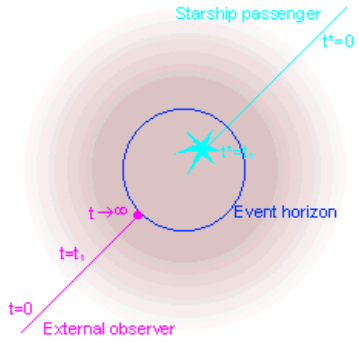



Fig. 1.17 Sequence of events as a person enters a black hole

The sequence of events as a starship passenger approaches the event horizon is illustrated in Fig. 1.17. The lower left corner depicts what an external observer would see, the upper right corner shows the perception of the person who falls into the black hole. Entering the event horizon, he would get a distorted view of the outside world at first. However, while falling towards the center, the starship and its passenger would be virtually stretched and finally torn apart because of the strong gravitational force gradient. An external observer outside the event horizon would perceive the starship and its passenger move slower the closer they get to the event horizon, corresponding to a time dilatation. Eventually, they would virtually seem to stand still exactly on the edge of the event horizon. He would never see them actually enter the black hole. By the way, this is also a reason why a black hole can never appear completely "black". Depending on its age, the black hole will still emit a certain amount of (red-shifted) radiation, aside from the Hawking radiation generated because of quantum fluctuations at its edge.

## 1.7 Examples of Relativistic Travel

**A trip to Proxima Centauri** As already mentioned in the introduction, it is essential to overcome the limitations of special relativity to allow sci-fi stories to take place in interstellar space. Otherwise the required travel times would exceed a person's lifespan by far.

> **Side note** Several of the equations and examples in this sub-chapter are taken from [Ger89].

Let us assume a starship with a very advanced, yet slower-than-light (STL) drive, were to reach Proxima Centauri, about 4ly away from Earth. This would impose an absolute lower limit of 4 years on the one-way travel. However, considering the drastic increase of mass as the ship approaches c, an enormous amount of energy would be necessary. Moreover, we have to take into account a limited engine power and the limited ability of humans to cope with excessive accelerations. A realistic STL travel to a nearby star system could work with a turn-around starship as shown in Fig. 1.13 which will be assumed in the following.

To describe the acceleration phase as observed from Earth's frame of reference, the simple relation v=gt for non-relativistic movements has to be modified as follows:

$$g \cdot t = \frac{v}{\sqrt{1-\frac{v^2}{c^2}}} \Rightarrow v = c \cdot \sqrt{\frac{(g \cdot t/c)^2}{1+(g \cdot t/c)^2}}$$

Eq. 1.20

> **Side note** It's not surprising that the above formula for the effective acceleration is also determined by the factor gamma, yet, it requires a separate derivation that I don't further explain to keep this chapter brief.

The relativistic and (hypothetical) non-relativistic speeds at a constant acceleration of $g=9.81ms^{-2}$ are plotted over time in Fig. 1.18. It would take 409 days to achieve 0.5c and 2509 days to achieve 0.99c at a constant acceleration of g. It is obviously not worth while extending the acceleration phase far beyond 0.5c where the curves begin to considerably diverge. It would consume six times the fuel to achieve 0.99c instead of 0.5c, considering that the engines would have to work at a constant power output all the time, while the benefit of covering a greater distance wouldn't be that significant.

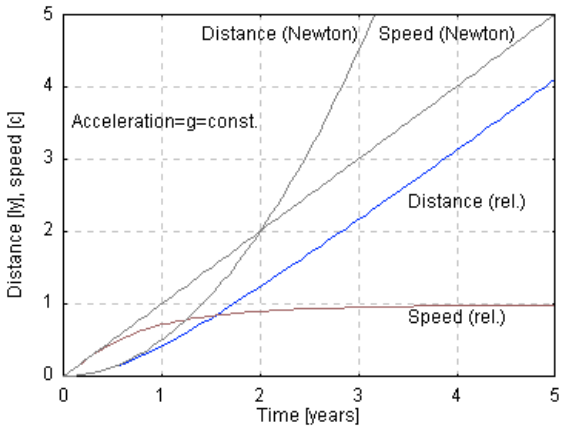



Fig. 1.18 Comparison of Newtonian and relativistic movement

To obtain the covered distance x after a certain time t, the speed v as given in Eq. 1.20 has to be integrated over time:

$$x = \int_0^t v(\tau)\,d\tau = c \cdot \int_0^t \sqrt{\frac{(g\cdot\tau/c)^2}{1+(g\cdot\tau/c)^2}}\,d\tau = \frac{c^2}{g}\left[\sqrt{1+(g\cdot\tau/c)^2}\right]_0^t = \frac{c^2}{g}\left[\sqrt{1+(g\cdot t/c)^2}-1\right]$$

Eq. 1.21

**Side note** Note that the variable tau instead of t is only used to keep the integration consistent and satisfy mathematicians ;-), since t couldn't denote both the variable and the constant.

There are two special cases which also become obvious in Fig. 1.18: For small speeds gt<<c Eq. 1.21 becomes the Newtonian formula for accelerated movement $x=\frac{1}{2}gt^2$. Therefore the two curves for non-relativistic and relativistic distances are almost identical during the first few months (and v<0.5c). If the theoretical non-relativistic speed gt exceeds c (which would be the case after several years of acceleration), the formula may be approximated with the simple linear relation x=ct, and the according graph is a straight line. This is evident, since we can assume the ship has actually reached a speed close to c, and the effective acceleration is marginal. A distance of one light-year would then be bridged in slightly more than a year.

If the acceleration is suspended at 0.5c after the aforementioned 409 days, the distance would be 4.84 trillion km which is 0.51ly. With a constant speed of 0.5c for another 2509-409=2100 days the ship would cover another 2.87ly, so the total distance would be 3.38ly. On the other hand, after additional 2100 days of acceleration to 0.99c our ship has bridged 56,5 trillion km, or 5.97ly. As we could expect, the constant acceleration to twice the speed is not as efficient as in the deep-sublight region where it should have doubled the covered distance.

A maximum speed of no more than 0.5c seems useful, at least for "close" destinations such as Proxima Centauri. With an acceleration of $g=9.81ms^{-2}$ the flight plan could look as follows:

| Flight to Proxima Centauri | Speed | Distance | Earth time | Ship time |
|---|---|---|---|---|
| Acceleration @ g | 0 to 0.5c | 0.51ly | 1.12 years | 0.96 years |
| Constant speed | 0.5c | 2.98ly | 5.96 years | 5.12 years |
| Deceleration @ -g | 0.5c to 0 | 0.51ly | 1.12 years | 0.96 years |
| **Total** | **-** | **4ly** | **8.20 years** | **7.04 years** |

Tab. 1.1 Flight plan for an STL trip to Proxima Centauri

The table already includes a correction of an "error" in the above considerations which referred to the time t as it elapses on Earth. The solution of the twin paradox revealed that the space traveler who leaves Earth at a certain speed and stops at the destination changes their frame of reference each time, no matter whether or not we take into account the effects of the acceleration phases. This is why the special relativistic time dilatation will become asymmetric. As the space traveler returns to Earth's frame of reference -either by returning to Earth or by landing on Proxima Centauri which can be supposed to be roughly in the same frame of reference as Earth- he will have aged less than his twin on Earth. During his flight his ship time t* elapses slower than the time t in Earth's frame of reference:

$$t = \frac{t^*}{\sqrt{1-\frac{v^2}{c^2}}} \Rightarrow t^* = t\cdot\sqrt{1-\frac{v^2}{c^2}}$$

Eq. 1.22

Eq. 1.22 is valid if the speed v is constant. During his constant speed period of 5.96 years in Earth's frame of reference the space traveler's clock would proceed by 5.12 years. In case of an acceleration or deceleration we have to switch to infinitesimal time periods dt and dt* and replace the constant velocity v with v(t) as in Eq. 1.20. This modified equation has to be integrated over the Earth time t to obtain t*:

$$t^* = \int_0^t \sqrt{1-\frac{v(\tau)^2}{c^2}}\,d\tau = \int_0^t \sqrt{1-\frac{(g\cdot\tau/c)^2}{1+(g\cdot\tau/c)^2}}\,d\tau = \int_0^t \frac{d\tau}{\sqrt{1+(g\cdot\tau/c)^2}} = \frac{c}{g}\int_0^t \frac{d\tau}{\sqrt{g/c+\tau^2}} = \frac{c}{g}\,\mathrm{ar\,sinh}\,\frac{g\cdot t}{c}$$

Eq. 1.23

**Side note** The function arsinh is called "area sinus hyperbolicus".

The space traveler would experience 0.96 years during the acceleration as well as the deceleration. The times are summarized in Tab. 1.1, yielding a total time experienced by the ship passenger of 7.04 years, as opposed to 8.20 years in the "resting" frame of reference on Earth.

**Traveling to the edge of the universe** The prospect of slowing down time as the ship approaches c offers fascinating possibilities of space travel even without FTL drive. The question is how far an STL starship could travel within a passenger's lifetime, assuming a constant acceleration of g=9.81ms$^{-2}$ all the time. Provided there is a starship with virtually unlimited fuel, the following theory would have to be proven: If the space traveler continued acceleration for many years, his speed would very slowly approach, but never exceed c, if observed from Earth. This wouldn't take him very far in his lifetime. However, according to Eq. 1.23 time slows down more and more, and this is the decisive effect. We might want to correct the above equations with the slower ship time t* instead of the Earth time t. We obtain the ship speed v* and distance x* if we apply Eq. 1.23 to Eq. 1.20 and Eq. 1.21, respectively.

$$v^* = c\cdot\sqrt{\frac{\sinh^2\frac{gt^*}{c}}{1+\sinh^2\frac{gt^*}{c}}} = c\cdot\tanh\frac{gt^*}{c}$$

Eq. 1.24

$$x^* = \frac{c^2}{g}\left[\sqrt{1+\sinh^2(gt^*/c)}-1\right] = \frac{c^2}{g}\left[\cosh(gt^*/c)-1\right]$$

Eq. 1.25

Note that the term tanh(gt*/c) is always smaller than 1, so that the measured speed always remains slower than c. On the other hand, x* may rise to literally astronomical values. Fig. 1.19 depicts the conjectural travel to the edge of the universe, roughly 10 billion light-years away, which could be accomplished in only 25 ship years! The traveler could even return to Earth which would require another 25 years; but there wouldn't probably be much left of Earth since the time elapsed in Earth's frame of reference would sum up to 10 billion years, obviously the same figure as the bridged distance in light-years.

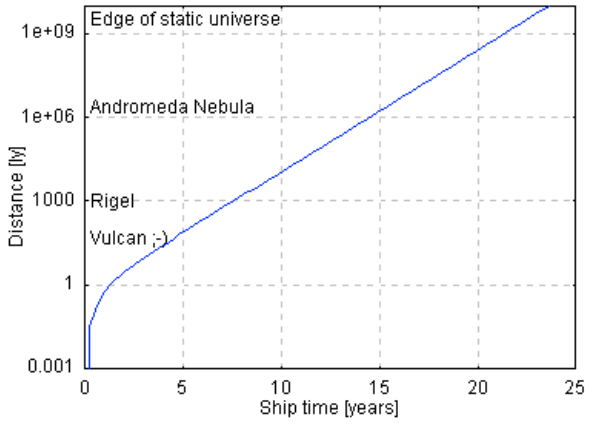



Fig. 1.19 Plan for an STL trip to the edge of the universe

**Side note** Apart from the objection that there would be hardly unlimited fuel for the travel, the above considerations assume a static universe. The real universe would further expand, and the traveler could never reach its edge which is probably moving at light speed.

**Fuel problems** The non-relativistic relation of thrust and speed was discussed in Section 1.5. If we take into account relativistic effects, we see that at a constant thrust the effective acceleration will continually decrease to zero as the speed approaches c. The simple relation v=gt is not valid anymore and has to be replaced with Eq. 1.20. Thus, we have to rewrite the fuel equation as follows:

$$\frac{v_1}{\sqrt{1-\frac{v_1^2}{c^2}}} = c\cdot\ln\frac{m_0}{m_1} \quad\Rightarrow\quad m_1 = m_0\cdot e^{-\frac{v_1}{c\sqrt{1-\frac{v_1^2}{c^2}}}}$$

Eq. 1.26

The two masses $m_0$ and $m_1$ still denote non-relativistic rest masses of the ship before and after the acceleration, respectively. Achieving $v_1$=0.5c would require not much more fuel than in the non-relativistic case, the payload could still be 56% of the total mass compared to 60%. This would be possible, provided that a matter/antimatter power source is available and the power conversion efficiency is 100%. If the aspired speed were 0.99c, the ship would have an unrealistic fuel share of 97%. The flight to the edge of the universe (24 ship years at a constant apparent acceleration of g) would require a fuel mass of 56 billion times the payload which is

beyond every reasonable limitations, of course.

If we assume that the ship first accelerates to 0.5 and then decelerates to zero on the flight to Proxima Centauri, we will get a still higher fuel share. Considering that Eq. 1.26 only describes the acceleration phase, the deceleration would have to start at a mass of $m_1$, and end at a still smaller mass of $m_2$. Taking into account both phases, we will easily see that the two mass factors have to be multiplied:

$$m_2 = m_1 \cdot e^{-\frac{v_1}{c\sqrt{1-\frac{v_1^2}{c^2}}}} = m_0 \cdot e^{-\frac{v_1}{c\sqrt{1-\frac{v_1^2}{c^2}}}} \cdot e^{-\frac{v_1}{c\sqrt{1-\frac{v_1^2}{c^2}}}}$$

Eq. 1.27

This would mean that the payload without refueling could be only 31% for $v_1$=0.5c. For $v_1$=0.99c the ship would consist of virtually nothing but fuel. Just for fun, flying to the edge of the universe and landing somewhere out there would need a fuel of $3*10^{21}$ tons, if the payload is one ton (Earth's mass: $6*10^{21}$t).

# 3 Subspace (Draft)

***The Physics and Technology of Warp Propulsion***

3.1 What is Subspace? - 3.2 Subspace Models - 3.3 Subspace Fields and Warp Fields - 3.4 Subspace Technology - 3.5 Subspace Phenomena

## 3.1 What is Subspace?

**Preliminary remark** Subspace as shown in Star Trek does not exist in present-day physics as a reality or only a theory. It is a concept of storytelling. Subspace is necessary because FTL travel is impossible within the boundaries of special relativity as outlined in Chapter 1.2. Being part of a fictional universe, subspace has not been developed by far as consequently as a truly scientific idea would have required. On the contrary, everything we know about subspace is based on more or less random and often contradictory canon evidence.

*The Physics and Technology of Warp Propulsion*

- Introduction
- Chapter 1 - Real Physics and Interstellar Travel
- Chapter 3 - Subspace
- Chapter 6 - Warp Speed Measurement
- Chapter 7 - Appendix

> **Important note** A Google search for "subspace" gives us more than 90% Trek-related results. Most of the rest refers to concepts where a subspace is a subset to a space in a mathematical sense. In other words, the real subspace is a generic mathematical model, whereas Trek's subspace denotes a specific physical concept. The equal naming is nothing more than a coincidence.

Much of the following is speculation and not strictly based on observations, but it is conceived to comply with them. Some prominent subspace characteristics and phenomena will be made more plausible. Still, some contradictory evidence may remain simply because subspace was never meant to represent a scientific theory.

**Definition** There is no canon definition of what subspace actually is. But there are countless canon accounts on subspace that allow to make a couple of presuppositions.

**1.** Subspace must be present everywhere in the universe because sensors can measure subspace distortions wherever deemed necessary. So it looks like for every point in our space there is a corresponding point in subspace, otherwise the various measurements of subspace stress in a certain region of space and of the warp field which obviously has a spatially limited effect on subspace would make no sense.

**2.** It appears that subspace is a continuum in addition to and outside of the three dimensions of our normal space and to time. Subspace may be described as a dimension, and there would be plenty of them left in present-day string theories that subspace could represent. But it is most likely not equivalent to a fourth spatial dimension, since it exhibits considerably different characteristics than the three familiar dimensions. Submerging into subspace is not like moving in normal space. This specialty of subspace is not unlike the dimension of time. Time is not equivalent to spatial coordinates either, since "time's arrow" dictates a preferred direction. Consequently, we would not expect subspace "distances" to be measured in meters (or even in seconds).

**3.** Transitions of matter or energy into subspace or, vice versa, from subspace into real space might occur naturally. Subspace itself or domains accessible only through subspace are even believed to be populated by indigenous lifeforms (see 3.5 Subspace Phenomena). It really exists as a physical domain and not just as a mathematical model illustrating the effects of FTL propulsion.

**4.** Still, the presence of subspace as depicted in Star Trek is beyond human perception as long as we are confined to our space-time, and it can only be verified with special measuring equipment of the future. The frequent light effects connected with subspace may be explained away in that they are secondary effects in our space, which are typical of, but not a definite indicator of the presence of a subspace phenomenon. In other words, present-day equipment wouldn't detect anything outlandish about subspace light effects, because it's just light. Only that we may not be able to explain where it comes from and why without the knowledge of subspace.

**Subspace vs. hyperspace** "Hyperspace" is another term frequently used in science fiction. It seems that most science fiction is using a concept of folded space for FTL travel, bringing two points in space closer together in a higher dimension -- hyperspace. This is much the same as the concept of wormholes which are commonplace in Star Trek too. It is quite obvious that subspace in Star Trek

is supposed to be something different - not only quantitatively (Star Trek ships with warp drive are considerably slower than if they were passing through a wormhole), but also qualitatively (ships at warp always remain practically in our space-time, whereas wormhole/hyperspace travelers definitely don't). Hyperspace was only mentioned very few times in Star Trek, and mostly by accident like a few times on TNG. We wouldn't expect the two domains to be identical, and they have different (even opposite) names. So where "hyperspace" was used like "subspace" on screen, we should either imagine that the characters actually said "subspace" (the convenient solution) or assume that they were really talking about something different, much less common than subspace.

## 3.2 Subspace Models

**Premise** We assume that subspace is a physical reality, an inhabitable realm and not just a mathematical model. Yet, in both cases we would have to describe how subspace could look like from the inside and how it could relate to real space. This chapter attempts to create a theoretical foundation for everything that we know about the structure of subspace from canon accounts.

> **Side note** Many of the considerations about the structure of subspace are based on Christian Rühl's excellent Subspace Manual at Star Trek Dimension [Rüh].

**Continuous vs. discrete model** Fig. 3.1 depicts a three-dimensional representation of the four-dimensional combination of space (x,y,z) and subspace (zeta). The model on the left is continuous, the one on the right exhibits distinct layers of subspace. In both models our three-dimensional space is just a part of a whole which we may call subspace. Only the two dimensions x and y of normal space are shown. The third axis with the coordinate zeta is what we can call the "subspace depth" which, as mentioned above, probably isn't measured in meters or other length units because subspace is not akin to a spatial dimension. In TNG: "Schisms" Geordi identified Riker's homing signal as coming from an energy level of 16.2 keV, which indicates that the scale of "subspace depth" may be an energy scale.

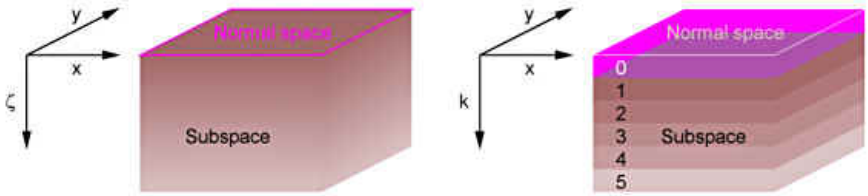



Fig. 3.1 Continuous (left) and discrete (right) subspace model

In the continuous model (Fig. 3.1 left) our normal space is comprised just of the very surface (x,y,zeta=0) of a subspace continuum. Zeta is continuous just like x and y too and may assume any positive real value, while zeta<0 remains undefined. Any point in the x-y plane has an infinite number of corresponding points in subspace along the zeta-axis. This is the model we are likely to apply as a first approach, assuming that x,y,z as well as time are continuous (at least to our perception and within the accuracy of conventional measuring equipment), and that there would be no reason why the same should not apply to subspace. But this model may turn out flawed.

The discrete model (Fig. 3.1 right) may work better to describe the expected as well as the observed properties of subspace. In this model, there is no continuous "subspace depth" zeta any longer. Instead of that, subspace is divided into discrete layers which I have given tentative numbers k starting with "0". Within each layer, the properties of subspace are changing only slightly or are even constant, while there is some sort of barrier between each two layers. We may imagine that considerably more energy has to be expended to move from one layer to the next one than inside the same layer. The bright green uppermost layer numbered with "0" is identical to normal space.

> **Side note** The concept of energy barriers is quite common in solid-state physics and particularly important to describe semiconductors. Any solitary atom exhibits discrete energy levels. In an extended solid-state material, these energy levels join to bands which are continuous in x,y,z and allow spatial charge transport if the band is not already fully occupied. There are band gaps which are not allowed to be occupied by charge carriers. In order to make the transition from one band to a higher (and emptier) one, it is necessary to supply a sufficient energy in the form of light, voltage or heat, corresponding to the width of the band gap.

The first problem with the continuous approach is that normal space would have to occupy a "thickness" of exactly zero. But realistically, there would have to be natural fluctuations ("zero-point energy"). Anyone or anything in our world would not always be exactly in real space, but slightly submerged into subspace. Perhaps this natural fluctuation is small enough to evade our senses, and it may have influenced any physical experiments prior to the discovery of subspace only below the measurement limit, so it remained unnoticed. Still, there is the problem that the subspace axis extends only in one direction, that there are supposed to be no negative values. As a result, the average of all fluctuations of normal space could not be found at zeta=0, but at a small subspace depth zeta>0. In other words, the actual normal space as found by measurements would not be equal to the theoretical normal space at the very surface of the continuum. It may be suggested that negative subspace is actually identical to hyperspace, but we would expect hyperspace to have completely different properties which may not be explained with a simple sign reversal. Summarizing, using a discrete model, it is much easier to explain why subspace is not within the limits of our perception and measuring equipment, the reason being that first a threshold has to be crossed to become aware of the existence of subspace. Some sort of exotic energy is required in considerable amounts to that end, energy which naturally exists because of quantum fluctuations only in minute amounts.

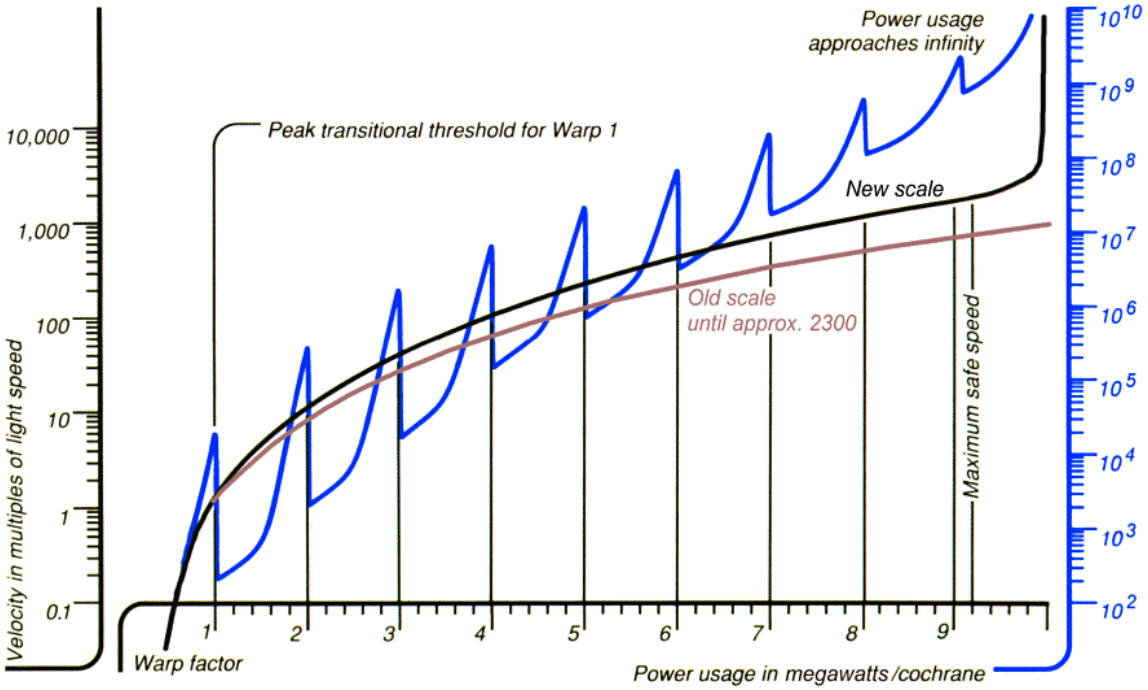



Fig. 3.2 Peak transitional thresholds

The second flaw of the continuous approach is more straightforward. We know that there are peak transitional thresholds between the familiar warp factors at which the energy expenditure rises strongly. More precisely, these thresholds even define the warp factors, as a look at the official diagram from the *TNG Technical Manual* [Ste91] reveals (part of it was clearly visible on screen in ENT: "First Flight" too). It seems logical that this observation is attributed to the structure of subspace. In other words, the peak transitional thresholds most likely exist due to transitions from one subspace layer to another. Alternatively, the peak transitional thresholds might reflect just the technical principle of the ship's engines, but in this case the the scale would be far less universal. It would change too often following progress in propulsion technology, and the warp factors would depend on the ship class, maybe even on how the drive is tuned on an individual ship.

In addition, we have evidence from TNG: "Schisms" that the aliens' inhabitable domain is located in a *"tertiary subspace manifold"*, which sounds very much like subspace consists of discrete layers. Geordi also mentions *"subspace bands"* in the episode, but just like radio frequency bands these may be defined more or less arbitrarily even in a continuous subspace model.

**Inside subspace** Having decided that subspace most likely has a discrete structure, it needs to be clarified what is the difference between the single subspace layers. The first assumption is that subspace layer 1 should be "somewhere else" than the normal space layer 0, as seen most obviously in TNG: "Schisms" where crew members were abducted into subspace by aliens. The realm shown in the episode looked much like our normal space, but was obviously not present in normal space at any place x,y,z or any time t (it was explicitly mentioned in the episode to be a *"tertiary subspace manifold/domain"*). Although we cannot be sure whether the alien realms in "Schisms" as well as in VOY: "Heroes and Demons" or VOY: "Bride of Chaotica" are not actually already in a parallel space/universe (see Fig. 3.4), there should be some inhabitable place "inside subspace" too if there is something like that on either side of it. In other words, if subspace is always talked about like an extended domain, why should it be only some sort of uninhabitable gate between the universes in reality?

There must be some sort of difference between the single layers. Regardless of the nature of this difference, we may imagine that subspace differs the more from normal space the further "below" the layer is located. We may say that subspace should become the more "exotic" the deeper we submerge into it. It is just speculation, but we may think of a model in which the subspace layers form some sort of funnel. In Fig. 3.1 we could see that each point in normal space-time (with just the x-y plane depicted) corresponds with one point at a given "depth" zeta (continuous) or k (discrete), still assuming that subspace has the same "lateral" geometry (in x-y) as normal space. Fig. 3.3 illustrates how a region in normal space could be related to gradually denser subspace layers. This compression could be the key to warp propulsion, transwarp and FTL communication, considering that the distances to be bridged in subspace in a given time shrink with respect to normal space! This way subspace can be a shortcut through space.

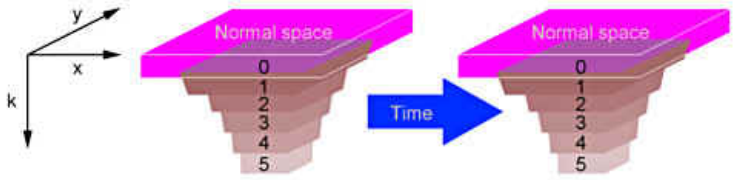



Fig. 3.3 Model of gradually compressed subspace

Unlike it is the case with wormholes or with some of the natural or artificial anomalies discussed below, these funnels are just illustrations of how subspace relates to normal space. They are no dedicated gates. Subspace is usually "flat" in a similar fashion as normal space-time is homogenous, at least in the absence of a mass that could distort it, according to general relativity. In a "flat" subspace one could draw such a funnel for each possible square unit in the x-y plane. Also, no starship entering subspace would "fall" into such a funnel. On the contrary, seeing how the layers are likely separated by energy barriers or peak transitional thresholds, they could much easier move within one plane than between the planes.

**The far end** We may go one step further and ponder where the bottom of subspace is located, and if the funnels ends up in a singularity, a depth in which all the reference points to our universe merge to one single point. In such a singularity space would be concentrated much like mass is concentrated in a single point in a mathematical model of a black hole. This is the point that Tom Paris would have arrived at in VOY: "Threshold" if we believe the fairy-tale of "infinite speed". But what if subspace actually doesn't reach down to such a singularity, if after passing a number of compressed layers its density drops again? This would be equivalent to approaching normal space again, yet on the "other side" of subspace. Passing through all subspace layers may take us to a parallel universe! This idea is illustrated in Fig. 3.4. It is undetermined how many parallel universes exist and which funnel leads to which universe. It may depend on when and how the parallel universe came to life. But basically it would be possible to have just one parallel universe like the one of TOS: "Mirror, Mirror", as well as an immense number like in TNG: "Parallels".

Evidence for the existence of parallel universes accessible through subspace can be found in DS9: "Playing God" where the protouniverse started off as a subspace phenomenon. The strange realms within subspace where fantasy becomes reality, like in TNG: "Where No One Has Gone Before" and "Remember Me", and perhaps also the aliens' inhabitable domain in "Schisms" may be well regarded as either deep in subspace or already beyond the far end of subspace.

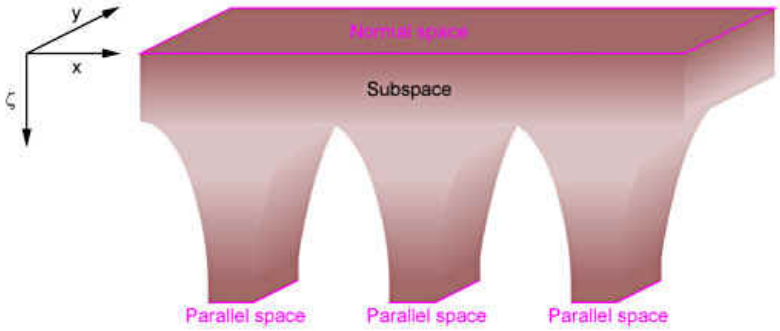



Fig. 3.4 Relative subspace model [Rüh]

Now we have to do just one more thing. We need to translate the "geocentric" model in Fig. 3.4 which arrogantly puts our universe above everything else to an absolute model in which our universe is just one among many others. This is shown in Fig. 3.5 where the irregularity is intentional to demonstrate that universes embedded into a subspace substrate may be regarded like galaxies inside the dimensions of space.

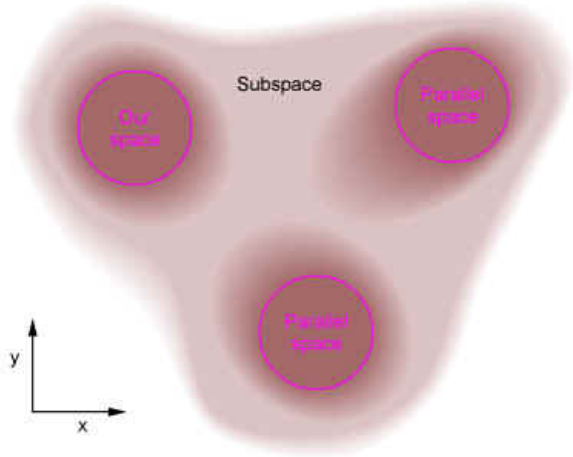



Fig. 3.5 Absolute subspace model [Rüh]

### 3.3 Subspace Fields and Warp Fields

**Premise** Subspace is the domain which enables warp propulsion, at least in the type of warp drive used on Federation starships and those of most known races in Star Trek. We know from many statements in episodes that the warp engine generates a warp field that totally encompasses the starship. The warp field corresponds to a distortion of subspace. The subspace field stress is the defining quantity of a subspace or warp field and is measured in cochranes. At a field stress of more than 1 cochrane we obtain a warp field which, with the right frequency and geometry tuning, may become a propulsive asymmetric warp field, as outlined in the *TNG Technical Manual* [Ste91]. The *TNGTM* describes such a warp field as a combination of different nested layers, *"each layer exerting a controlled amount of force against its next-outermost neighbor."* This coupling of nested layers which correspond to different frames of reference can be supposed to enable FTL speeds because between each two layers the effective relative speed would remain below c. Still, with no more than 9 layers the attainable speed could be no more than 9c. Star Trek's warp drive circumvents the limitations of special relativity, but there must be something special about subspace as a frame of reference in addition. As outlined in the previous chapter, one idea is that the structure of subspace may be such that corresponding points in space are closer together in subspace.

> **Side note** A more elaborate and in several ways different treatise on the nature of subspace fields is Subspace Physics [Hin2] at Jason Hinson's site

In order to achieve superluminal speeds, would the ship itself have to submerge into subspace and hence into a separate domain, like

it is with (hyperspace) jump gates or with wormholes? Not necessarily. The term "subspace field" itself, as well as the fact that the ship is not visibly in a different realm during normal warp travel indicate that it may still be in normal space. Conversely, with the various transwarp drives or slipstream that could be seen especially during TNG and Voyager the ship always appeared to be in some sort of tunnel through space which would have to be similar with normal warp propulsion if it took place inside subspace. Additionally we know from TNG: "Where No One Has Gone Before", "Schisms" and other occasions that subspace is a realm that is inhabitable but that ought to look in some way different than normal space. Hence, it seems more likely that the a ship using normal warp drive remains in our space. In this sense a subspace field is not a field that opens a gate into subspace, but rather one that allows subspace to "flow" into our space, making it more like subspace or shielding it against our space like some sort of semi-permeable wall.

**Subspace field** A subspace field below 1 cochrane is a distortion of normal space for the most part, so we may want to call it just a "space field" (but since subspace includes space, the collective name "subspace field" is definitely correct). The field is not strong enough to stretch down to the first layer of subspace. But it has the effect of lowering the apparent mass of the ship as observed from outside because of its special geometry. This effect of a subspace field would be much the same if the ship were amidst the distortion, but the stress would kill the crew and ultimately destroy the ship at above a few hundred millicochranes. Instead of exposing the crew and spaceframe to such a stress, all subspace fields, also and especially those above 1 cochrane, are created in a way that a shell of high distortion envelops the ship which is safe in a largely undistorted interior, as can be seen in Fig. 3.6.

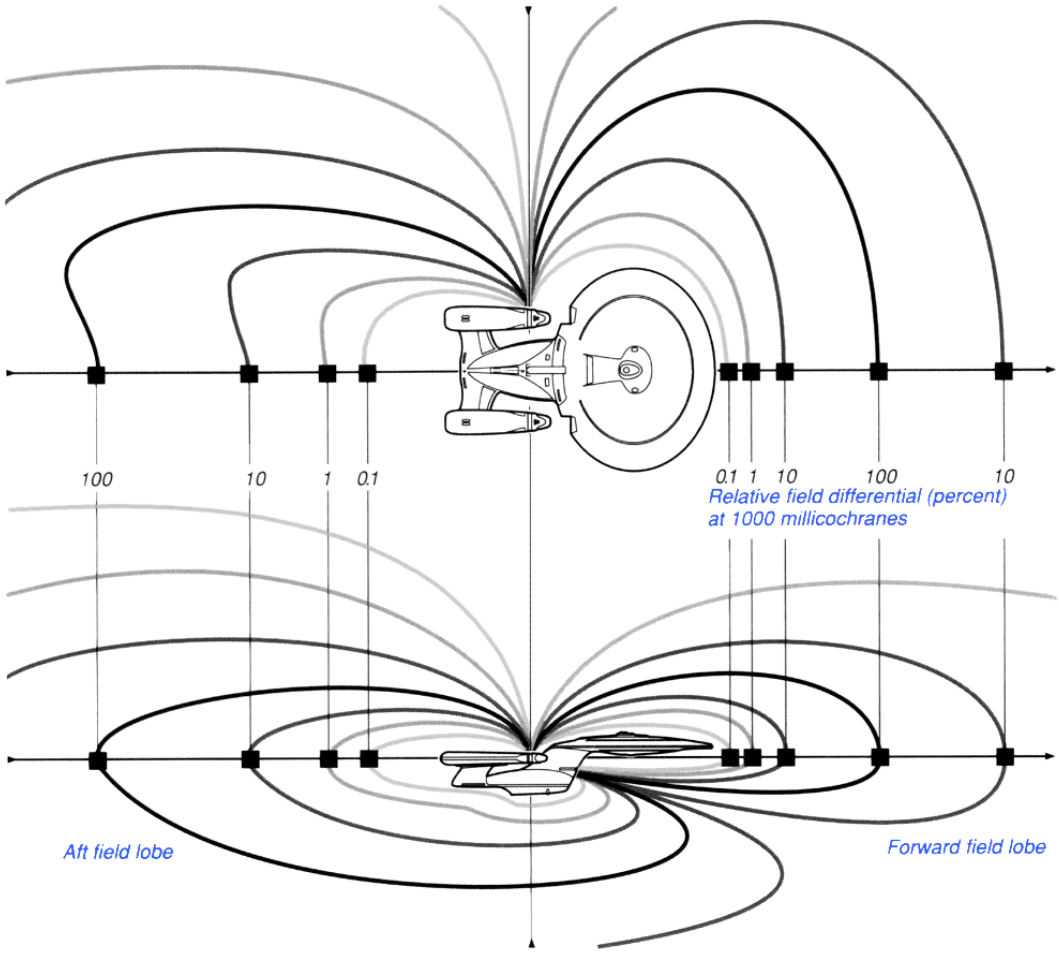



Fig. 3.6 Warp field geometry

The mass-lowering property of subspace fields is used in the impulse engines to facilitate sublight propulsion through subspace driver coils in the impulse engines as stated in the *TNGTM*. Once back in normal space, the exhausted plasma from the impulse engines gain mass and because of the principle of momentum conservation the thrust increases accordingly -- the ship may have less powerful impulse engines and needs to carry less fuel. For warp, on the other hand, it may be just a beneficial side effect. The mass would have to be lowered to zero to be able to reach light speed, which is impossible with any field of limited energy content, even a warp field of above 1 cochrane.

Fig. 3.2 shows the power expenditure of a typical warp drive. If we look at the first peak transitional threshold at Warp 1, we find that the power rises strongly as the speed approaches the particular threshold which is exactly at the speed of light, c. The power peak is almost two decades high, but clearly not infinite. If, however, the mass can't be lowered to zero, how can the threshold be passed with finite energy? The reason must lie in some limited interaction of the subspace field below 1 cochrane with the first subspace layer. While there may be no open gate that would allow energy to flow into subspace or subspace to flow into normal space, some tunneling may occur between the two, considering that the space distortion lowers the threshold between the two domains. Any subspace or warp field does the same in any other layer of subspace.

**Symmetric warp field** A symmetric warp field is a subspace field strong enough to shield a ship from normal space in a way that it takes on the frame of reference of subspace, while the ship itself is still in normal space. In order to become a warp field, the subspace stress has to be more than 1 cochrane. At 1 cochrane the field created by the subspace coils in normal space is strong enough to cross the threshold to the first subspace layer. Vice versa, subspace begins to flow into normal space. This subspace begins to "fill" the spatial distortion shell surrounding the ship, effectively shielding it against our universe. If a symmetric warp field is

created around a ship, it will remain in our universe, but it will be enabled to move as if it were in subspace. Only that a symmetric field itself does not move the vessel. It necessitates either an additional engine to achieve propulsion or, more elegantly, the warp field has to be tuned to become asymmetric and thereby propulsive.

We may surmise that many alien warp ships without distinct nacelles, especially those with rocket engines at the aft end, use symmetric warp fields and hence need the additional engines to push the ship forward. Federation starships, on the other hand, have additional engines just for sublight propulsion, while at warp only the warp nacelles are activated, creating a propulsive asymmetric warp field.

Symmetric warp fields are in use in the computer cores of modern Federation starships as we know from the *TNGTM*. With all of the devices and wires of the computer core embedded into a symmetric warp field, it is possible for the charge carriers inside to accelerate to FTL speeds. At the interfaces to external components the current slows down and is limited to light speed.

**Asymmetric warp field** Asymmetric warp fields combine the mass-lowering and shielding properties of a symmetric warp field with a propulsive effect. This may achieved through creating a warp field by firing the coils in the nacelles such that the extending fields created by the single coils keep pushing on each other and are driven towards the aft end of the ship. This is about the explanation given in the *TNGTM*. The book, however, does not distinguish between multilayered fields that are needed for propulsion in the first place (because only through interaction between the partial fields or rather waves created by the coils the sum wave can have a preferred direction) on one hand and multilayered fields to achieve higher warp speeds on the other hand. The difference will become obvious in the following paragraph.

**Multilayered warp field** In order to be propulsive at all, the single field components generated by the warp coils must interfere or in some other fashion interact to create an imbalance between the fore and the aft direction. Hence, such fields would have to be of the same type, otherwise they could penetrate one another without such an interaction. This is why we can assume that one single frequency of energy is basically sufficient to drive the ship. Since frequency corresponds to energy and energy corresponds to subspace levels, we may speculate that the higher the frequency is, the deeper may the field extend into subspace. With higher energy and/or higher frequency the starship may achieve higher warp speeds.

It may be possible to achieve and maintain Warp 2, for instance, with the warp field still tuned for Warp 1 if only enough power is released into subspace to cross the Warp 2 threshold too. Looking at Fig. 3.2 again, we could extrapolate the curve piece below Warp 1, for instance, to estimate which power would be necessary to achieve Warp 2 with only one layer of the warp field. This would not really work out quantitatively, because in that case the power usage would be many orders of magnitude higher which is definitely a stretch because we wouldn't expect the subspace layers to be that much different in whatever property matters for propulsion. But qualitatively it may give us the reason why the frequency is tuned first of all to reach a certain threshold and the warp factor assigned to it. We can find an analogy in the real world in automobiles. Theoretically it would be possible to accelerate up to 100km/h in the first gear. But no one would do that because, aside from being highly inefficient, the engine and gear would not survive that for long. It may be advantageous to create nested multilayered warp fields instead of just one for the highest threshold because of the barrier lowering, with each warp field paving the way for the next higher field.

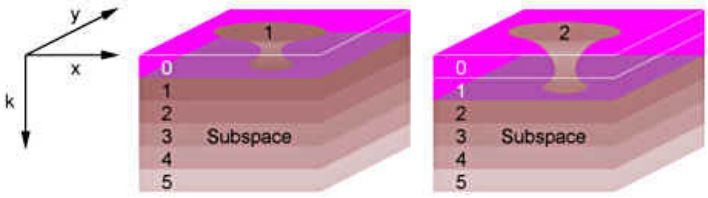



Fig. 3.7 Examples of warp fields: Warp 1 (left), Warp 2 (right)

**Warp factors** Fig. 3.7 depicts subspace fields of two strengths whose difference lies in how far they reach down into subspace. On the left image we can see that the field causes a transition of subspace layer 1 which flows into normal space. The right image depicts a stronger subspace field which corresponds to layer 2. We can easily correlate the depictions to cochrane values (TNG scale) and warp factors. On the left side we can see that the field has just exceeded the first threshold from normal space with index 0 to the subspace layer 1. It is a field of more than 1 cochrane and less than 10 cochranes, corresponding to speeds between Warp 1 and Warp 2. The field on the right would have between 10 and 39 cochranes or Warp 2 to Warp 3. Fig. 3.7 does not tell us anything about the structure of a warp field which, as was mentioned above, usually consists of not a single but of nested layers. But we may imagine that the sandwich structure of subspace is reflected by the structure of warp field layers, the reason being that their cumulative interaction allows to exceed the higher thresholds (the ones deeper in subspace).

> **The Planck time** In addition to the effect of coupled layers of subspace fields the *TNGTM* mentions one more property of subspace fields that may or may not be crucial for the working principle of the conventional warp drive. While at warp, the ship appears to be at superluminal speeds for less than the Planck time of $1.3*10^{-43}$s to the outside world. This suggestion may originally come from Jesco von Puttkamer, NASA director and scientific advisor of the first feature film, "Star Trek: The Motion Picture". It seems that the mention in the *TNGTM* was included rather as a homage to that first scientific attempt of explanation of warp drive, because the Planck time is not needed once we postulate that actually the different frame of reference of subspace allows FTL speeds. Well, unless subspace were just a model for the effect conceived by von Puttkamer. The *TNGTM* creates the impression that Cochrane's original drive principle, the "continuum distortion propulsion" (CDP) solely relied on the Planck time effect, rather than on subspace fields as they are used in later designs. But there is nothing in canon Trek to support this notion.

**Energy conservation** There is still one important property of warp drive still to be discussed. Warp propulsion requires a constant power output to maintain a certain speed or Warp factor. Once the power system fails, the ship will drop out of warp and slow down to sublight speed (a speed which will remain constant without acceleration because in motion in absence of a subspace field is governed by Newton's laws).

One theory popular in fandom is that a so-called "continuum drag" is the subspace equivalent to friction in classical mechanics -- a force that constantly counteracts the acceleration force of the ship and that increases with its speed. The continuum drag would be responsible for slowing the ship down to sublight speeds once the warp engine fails. Once at sublight speed, there would be no continuum drag any longer, and the friction with interstellar particles would be usually almost negligible.

**Subspace submersion and transwarp** Fig. 3.8 illustrates a different kind of warp field, one that allows a starship to enter subspace entirely, without being surrounded by normal space. In Fig. 3.9 two of such funnels are combined inside subspace to a tunnel. It is obvious that once in subspace a starship can move freely within this domain, without the need to drag a subspace field along that extends from normal space into subspace. It is possible that the power expenditure to maintain a certain speed inside subspace is lower. Only crossing one or more thresholds to submerge into subspace may consume considerably more energy than creating a normal warp field.

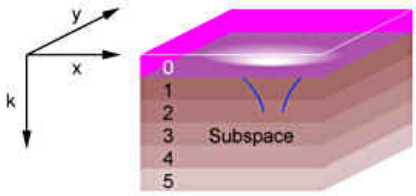



Fig. 3.8 Subspace submersion

Subspace submersion is employed in subspace communications. The *TNGTM* mentions that signals are pushed into deep subspace where they can propagate significantly faster. But without amplification they will surface after some 20 light-years and thus slow down to light speed. We can easily imagine that this effect is attributed to the "continuum drag". Transwarp of the type used by the Borg seems to be based on the same principle. The need to use pre-existing channels like first mentioned in TNG: "Descent" and later shown in VOY: "Endgame" corresponds with the high energy to push or drag the whole ship into deeper subspace layers.

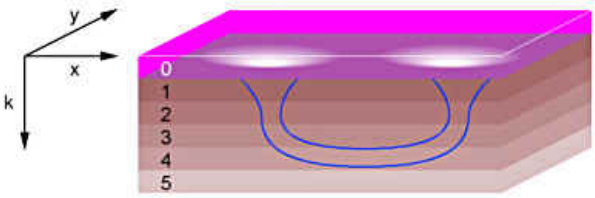



Fig. 3.9 Subspace channel

### 3.4 Subspace Technology

Not yet available.

### 3.5 Subspace Phenomena

Not yet available.

### More about Subspace

- Power & Propulsion - about the right intermix ratio, warp inside a star system, how to stop a starship etc.

# 6 Warp Speed Measurement

***The Physics and Technology of Warp Propulsion***

6.1 Concept of Warp Factors - 6.2 TOS Scale (23rd Century) - 6.3 TNG Scale (24th Century) - 6.4 Possible Future Scale

### 6.1 Concept of Warp Factors

**Definition** A warp factor is a unitless figure that represents the speed of a starship or of a signal traveling faster than light. There is a non-linear dependence between the warp factor and the effective FTL speed. In every known variant of the warp scale the speed rises exponentially with the warp factor, meaning that from Warp 1 to Warp 2 the speed more than doubles. The exponent is subject to vary between the scales.

There is no common canon symbol or abbreviation for warp factors, although "WF" or "wf" are sometimes used in textbooks. In spoken language they are referred to as *"Warp X"* (written with a capital "W") or, now rather antiquated, *"warp factor X"* during the time of TOS.

*The Physics and Technology of Warp Propulsion*

- Introduction
- Chapter 1 - Real Physics and Interstellar Travel
- Chapter 3 - Subspace
- Chapter 6 - Warp Speed Measurement
- Chapter 7 - Appendix

In all known warp scales "Warp 1" corresponds to the speed of light. The warp scale is continuous, meaning that real numbers such as "Warp 8.179" are possible, although it seems that starships most often travel at integer warp factors 7, 8, 9, etc.

Warp factors below Warp 1 are occasionally mentioned in Star Trek, mostly in the scope of the 24th century scale. Consequentially these refer to sublight speeds. The question whether something like Warp 0.5 exists has some relevance. Warp 0.5 could either mean that the warp drive also operates at sublight speeds, without a need to activate the impulse engines, or that the warp scale is simply extrapolated below Warp 1, even if only the impulse drive is on. With the note in the *TNG Technical Manual* [Ste91] that the impulse drive makes use of subspace driver coils the latter makes sense also technically, because if the warp factor generally describes the formation of a subspace field, it may as well apply to the field generated by the driver coils at sublight speed.

**Background** Bearing in mind that the speed of light c is already as high as 3*10^8m/s, it is obvious that with conventional units such as meters per second or kilometers per hour as they are used for the speeds of today's planes or spaceships we would end up with unreasonably large and unhandy figures when describing FTL motion. Generally, it may have been possible to switch to something like light years per day or similar manageable time and distance units. The reason why Starfleet introduced the apparently abstract and strangely non-linear warp factor must lie in the physical principle and/or the technical implementation of the propulsion system. In other words, the function of speed vs. warp factor most likely reflects how strong a warp field must be to achieve a certain speed.

> **Side note** Gene Roddenberry may have originally conceived warp factors for the sake of the dramatic impression of Kirk's commands. "Mr. Sulu, ahead Warp factor two." is more precise than "half speed ahead", and it sounds much more to the point than "increase to three light years per day". Moreover, with warp factors obscuring the real speed of the ship, calculation errors between speeds, times and distances mentioned on screen were less likely to occur. It was as late as in TNG that a correlation between all of the three figures was routinely made on screen (most often by Data), revealing the "true" significance of warp factors.

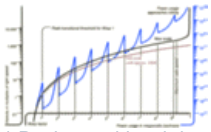


Fig. 6.1 Peak transitional thresholds

Fig. 6.2 Warp curve in ENT: "First Flight"

The sawtooth curve of the power expenditure vs. warp factor as shown in Fig. 6.1 and discussed in 3.3 Subspace Fields and Warp Fields is canon since it appeared on screen in ENT: "First Flight". The tips of the sawtooth are the so-called peak transitional thresholds, the points at which the power expenditure rises steeply because there is supposedly a physical threshold (in one possible interpretation, the transition to the next subspace layer) to be crossed. These peak transitional thresholds are located at the integer warp factors 1, 2, 3, and so on. Their existence is evidence that something significant happens at integer warp factors, justifying that the scale is based on these prominent points. In other words, the scale is designed in a way that the warp factors count up certain events that occur as the warp field is being expanded and intensified, which corresponds with an increase in speed.

Fig. 6.1 also shows a reason why ships customarily travel at integer warp factors or slightly above but never slightly below. We have to bear in mind that warp propulsion is non-Newtonian and a constant power supply is required to maintain a constant speed. If we imagine for a moment that the peaks are smoothed out, the power consumption (yellow sawtooth curve) overall increases stronger with the warp factor than the speed v/c (white for the TNG scale) does. In other words, the faster a ship goes, the higher is the energy consumption for the trip.

However, as we take into account the peak transitional thresholds, the power consumption per cochrane rapidly drops to a much lower value on the right side of each peak. More precisely, the drop happens slightly above the integer warp factor, but for the sake of simplicity we may assume that "increase speed to Warp 6" means to cross the sixth threshold, after which the ship may be actually at Warp 6.01. Here the power consumption is a lot lower than at Warp 5.99. It is obvious that staying slightly below Warp 6 would be very disadvantageous.

We can also see that a bit above Warp 6 the power consumption per cochrane is the same as at Warp 5.6. Since the cochrane value is equivalent to v/c, the effective increase of power consumption between Warp 5.6 and Warp 6.01 is the same as the speed increase. So the total energy consumption for a trip at Warp 6.01 is the same as at Warp 5.6 but the ship is some 25% faster. Hence, Warp 5.6 is some sort of break-even point, above which it is advantageous to speed up even further to Warp 6.01.

## 6.2 TOS Scale (23rd Century)

**Definition** During the era of The Original Series (TOS) a warp scale was in use where the warp factor was described with:

$$WF = \sqrt[3]{\frac{v}{c}}$$

Eq. 6.1

So the warp factor equals the cubic root of the ship's achieved speed ratio v/c, with v being the effective ship speed and c being the speed of light. Warp 1 corresponds to the speed of light.

> **Side note** The TOS scale is often referred to as "Cochrane scale" in fandom, but this is conjecture and, like the underlying formula, was never mentioned in the series.

Strictly speaking, the warp scale used in the TOS era is conjectural. There is no reference that would allow us to relate a warp factor stated on screen to a distance and a time. The now widely accepted scale, however, must have been established behind the scenes in a kind of writer's guide. David Whitfield, who had access to essential documents of Roddenberry's staff mentions the scale in his book *The Making of Star Trek* [Whi68] as soon as 1968. In his reproduction of the warp scale, Warp 1 denotes the speed of light, Warp 3 is 24c, Warp 6 is 216c and Warp 8 is 512c. Actually Warp 3 should be 27c, but with the other values following exactly the third-power law of Eq. 6.1 we have a confirmation that the scale was really being used at least as a guideline.

**Range** The continuous curve of the TOS scale (teal in Fig. 6.1) insinuates that warp factors below 1 and fractional values were well possible in this measurement system. Warp factors below 1, however, were not mentioned once in TOS, and fractions occurred only

occasionally. As late as in "Star Trek: The Motion Picture" (TMP) warp factors lower than Warp 1 appeared for the first time, when Sulu carefully activated the untested warp drive and slowly accelerated to warp speed. More precisely, the ship was already at Warp 0.5 but was only running on impulse drive when the warp drive was engaged. So it appears that impulse speeds can be measured with warp factors as well, the scale is uninterrupted. Since we shouldn't assume that the measurement system has changed between TOS and TMP, we can assume that warp factors below 1 already existed in TOS. In TOS, it often seemed like Kirk ordered to go to warp, a switch was flipped and the ship's speed immediately jumped up to Warp 1, but perhaps we should ascribe this immediate effect to dramatic license.

Warp factors higher than 10 are possible in the TOS warp scale. We know of a few incidents where the Enterprise was said to have achieved such high speeds that are well beyond the nominal top speed of Warp 8. The first one is in TOS: "The Changeling", where Nomad improved the Enterprise's engines to achieve speeds of up to Warp 11. Another one is in TOS: "That Which Survives", where a modification to the engines by Losira accelerates the Enterprise to as much as Warp 14.1. This is also one of very few examples of fractional warp factors. On another occasion, in TOS: "By Any Other Name", the Enterprise was tampered with by the Kelvans crossed the Galactic Barrier at Warp 11. The Kelvans were going to reach the Andromeda Galaxy, 2.3 million light years away, in three centuries. This gives us an indirect speed reference of as much as Warp 19.7. Yet, a definite figure was not mentioned for the warp factor during the long trip.

> There are two mentions of ultra-high warp speeds in TAS, both in "The Counter-Clock Incident". The Enterprise's sensors registered Karla Five's vessel with a speed of approx. Warp 36. The Enterprise herself was accelerated to speeds in the excess of Warp 22, due to a tractor beam link with the other ship.

**Background** As already mentioned above in Section 6.1, the warp factors likely reflect the physical principle and/or the technical implementation of the warp drive. The peak transitional thresholds and the corresponding prevalence of integer warp factors is a reason why the structure of the warp field and/or of subspace is most likely a layered one, as outlined in 3.3 Subspace Fields and Warp Fields. So in the TOS warp scale "Warp 1" may simply mean that the subspace field extends just into the first layer of subspace, having crossed the first peak transitional threshold. This may go on for Warp 2, 3, and so on. In the TOS scale there is no theoretical limit for warp factors and for the according submersion of the warp field into subspace. Only the power available on a ship imposes a limit to crossing the peak transitional thresholds of always deeper layers.

In TOS: "That Which Survived" one of the following two things may have happened when the ship attained Warp 14.1. The first possibility is that actually 14 subspace layers were subsequently crossed by tuning the warp field accordingly, although this has never been attempted before. The other possibility is that the Enterprise's engines were never built to penetrate more than eight layers (Warp 8 was said to be the maximum safe speed), and that simply so much power was pushed into the first to the eighth layer, overstressing these layers, that the equivalent warp factor was 14.1. In other words, in the latter case the Enterprise's instruments would not have registered further thresholds and may have been off scale anyway, so the equivalent warp factor would have been calculated from the actual speed.

**Use in Star Trek Enterprise** For the sake of plausibility, we should assume that the 22nd century (Star Trek Enterprise) uses the same warp scale as in the 23rd century (TOS), rather than the one of the 24th century (TNG). Enterprise NX-01 has a top speed of Warp 5, which in the TOS scale is reasonably slower than the max. speed of the Enterprise NCC-1701 (Warp 6) or its max. emergency speed (Warp 8). If Enterprise used the TNG scale, NX-01 would be almost as fast as the NCC-1701.

**Problems** There are some cases in which the original Enterprise must have traveled at extremely high speeds, even without alien modifications to the engines. The first example is Pike's line in "The Cage" that his ship comes from *"a stellar group at the other end of this galaxy"*. Also, in TOS the Enterprise crosses the Galactic Barrier no less than three times, which is supposedly at the very least a thousand light-years from Federation space. Finally, in "Star Trek: The Final Frontier" the Enterprise-A explicitly travels to the center of the galaxy, which is some 25,000 light-years away, a journey that would take a lifetime at Warp 8. Enterprise NX-01 too is occasionally much faster than Eq. 6.1 would allow. The most blatant case is in ENT: "Broken Bow" when the the travel time to Qo'noS is said to be *"Four days there, four days back."*, which would place the Klingon homeworld only 1 light-year away from Earth, considering that Enterprise's top speed at the time is merely Warp 4.5.

There are occurrences in other Trek series too where the starships are effectively many times faster than they should be, or travel to destinations that should be far out of reach. See also the examples in Section 6.3. These are most often fundamental plausibility problems of the stories, rather than of the warp speed measurement. Yet, it has been suggested that the actual speed of a starship at a given warp factor depends on environmental conditions.

The *TNG Technical Manual* [Ste91] states: *"The actual values [of the cochrane value and hence of the speed attainable at a warp factor] are dependent upon interstellar conditions, e.g., gas density, electric and magnetic fields within the different regions of the Milky Way galaxy, and fluctuations in the subspace domain. Starships routinely travel at multiples of c, but they suffer from energy penalties resulting from quantum drag forces and motive power oscillation inefficiencies."* In order to explain away warp speed inconsistencies along these lines, we need a correction factor k to give us the actual relation between warp factor and speed, modifying Eq. 6.1 for the TOS scale to WF=k*(v/c)^(1/3). While some regions of space may allow only below average speeds, most importantly there have to be other regions in which a much higher speed is attainable at a certain warp factor (known as "warp highways" in fandom). In these regions k has to be considerably lower than 1 and in extreme cases (travel across the whole galaxy) close to 0. There are a couple of problems with this assumption, however:

- If there is a coefficient k(x,y,z) of spatial conditions that locally influences the warp speed at a given warp factor, this k ought to play a role in the stories. It is totally implausible that it was never mentioned in any form. Well, we know of the low-warp Hekaras Sector in TNG: "Force of Nature", but here the problem is to maintain a stable warp field in the first place; there is no hint that in this region warp factors may correspond to lower speeds.
- The whole series Star Trek Voyager makes no sense under the assumption that there are warp highways everywhere. The crew was seeking wormholes and other rare phenomena all the time, not on-ramps and exits of warp highways. Well, thanks to Seven's work in the astrometric lab five years could be shaved off as mentioned n VOY: "Year of Hell, Part I", but on the display it looks like Seven, with the more accurate Borg technology, just plotted a straighter course than the previous one (or one with fewer obstacles to be avoided).
- A ship's captain (sea or space) orders a course and a speed. If a captain has to take into account the gradient of an additional coefficient along the flight path that strongly influences the speed, he either has to be a genius in mental calculus, or he doesn't really care how soon the ship actually arrives. Well, there is always the possibility of asking the computer to calculate the course and the travel time in

advance, but in Star Trek offhand ETA calculations and swift orders to go to a certain course and warp factor are commonplace. These would be pointless if the travel time to the same destination could vary between minutes and months, and the best course would not be a straight line.

- Well, we may expect that realistically the interstellar conditions don't vary too strongly, with a factor of perhaps 2 between the highest and the lowest possible speed at a given warp factor, which would alleviate the preceding point. Moreover, on very long journeys the varying conditions in the flight path will average out. But then the theory of fluctuating conditions couldn't explain phenomena such as the Enterprise's journey to the center of the galaxy at all.
- Eq. 6.1 and Eq. 6.2 give us quite simple relations between warp factor and speed, without a correction factor (or with k=1) or any other coefficients in it. If anything, these formulae look like they describe a "basic" or "optimum" condition of space (without interstellar dust, without interfering fields, etc.), rather than an average. And if they really describe an average as hinted at in the *TNGTM*, there is the question of the averaging method, the size of the sample, the sampling region and how frequently such an average would need to be updated.

Summarizing, there may be regions in which warp travel is facilitated, meaning that less power is required to achieve and/or maintain a certain warp factor. There may be other regions where the contrary is true (such as the Hekaras Corridor). But while the peak transitional thresholds may be lowered, they rather wouldn't shift with the natural inhomogeneity of space, because this would make the concept of warp factors pretty useless. And even if they do shift, it may not suffice to explain anomalous warp speeds.

### 6.3 TNG Scale (24th Century)

**Definition up to Warp 9** Some time between the TOS Movie era (late 23rd century) and the TNG era (mid-24th century) the warp scale changed. The new scale is valid in all series set in the 24th century (TNG, DS9, Voyager). In this new TNG scale Warp 1 still denotes the speed of light. But above Warp 1 the speed corresponding to a certain warp factor is generally higher than in the TOS scale. The discrepancy slowly grows between Warp 1, where the curves intersect, to Warp 9, where the speed is 729c according to the TOS scale but 1516c for the TNG one (see Fig. 6.1). The exact equation for speeds up to Warp 9 in the TNG scale is:

$$WF = \sqrt[3.\overline{3}]{\frac{v}{c}}$$

Eq. 6.2

So the exponent increased from 3 to 10/3=3.333..., accounting for the higher speed at a given warp factor.

**Definition between Warp 9 and 10** The difference between TOS and TNG scale, however, becomes dramatic above Warp 9. The warp speed table was changed in accordance with Roddenberry's wish that Warp 10 should be the absolute maximum speed.

> **Side note** The *TNG Technical Manual* [Ste91] states about the real-life rationale for the recalibration: *"Figuring out how 'fast' various warp speeds are was pretty complicated, but not just from a 'scientific' viewpoint. First, we had to satisfy the general fan expectation that the new ship was significantly faster than the original. Second, we had to work with Gene's recalibration, which put Warp 10 at the absolute top of the scale. These first two constraints are fairly simple, but we quickly discovered that it was easy to make warp speeds TOO fast. Beyond a certain speed, we found that the ship would be able to cross the entire galaxy within a matter of just a few months. (Having the ship too fast would make the galaxy too small a place for the Star Trek format.) Finally, we had to provide some loophole for various powerful aliens like Q, who have a knack for tossing the ship millions of light-years in the time of a commercial break. Our solution was to redraw the warp curve so that the exponent of the warp factor increases gradually, then sharply as you approach Warp 10. At Warp 10, the exponent (and the speed) would be infinite, so you could never reach this value. (Mike used an Excel spreadsheet to calculate the speeds and times.) This lets Q and his friends have fun in the 9.9999+ range, but also lets our ship travel slowly enough to keep the galaxy a big place, and meets the other criteria. (By the way, we estimate that in 'Where No One Has Gone Before' the Traveler was probably propelling the Enterprise at about Warp 9.9999999996. Good thing they were in the carpool lane.)"*

In the TOS scale the speed continued to increase with the third power as in Eq. 6.1, with Warp 10 being as "slow" as 1000c. In the TNG scale, beyond Warp 9, the exponent rises in a way that at exactly Warp 10 the speed (as well as the power expenditure) becomes infinite. In other words, the TNG scale ends at Warp 10. The recalibration compresses the whole range from about Warp 9 to infinite Warp factor in the TOS scale into the Warp 9..10 interval of the TNG scale.

A table in the *Star Trek Encyclopedia III* [Oku99] lists selected TNG warp factors and the corresponding speeds above Warp 9. But these figures don't seem to originate in Mike Okuda's Excel sheet mentioned in the *TNGTM*. In an e-mail from January 1995 Mike Okuda states, *"Between 9 and 10, I gradually increased the exponent so that it approached infinity as the warp factor approached 10. Lacking knowledge of calculus, I just drew what looked to me to be a credible curve on graph paper, then pulled the points from there."* So there does not seem to be an "official" underlying formula for the range between Warp 9 and Warp 10.

> **Side note** There are some fan-made approximations, notably the formula by Graham Kennedy at DITL and the ones on Joshua Bell's site for TNG warp factors beyond 9.

The following table compares the two warp scales:

| Warp factor | TOS scale | | TNG scale | | |
|---|---|---|---|---|---|
| | Rel. speed (v/c) | Formula | Rel. speed (v/c) | Formula | Notes |
| 0.5 | 0.125 | | 0.099 | | |
| 1 | 1 | | 1 | | Speed of light |
| 2 | 8 | | 10.08 | | |
| 3 | 27 | | 38.94 | | |
| 4 | 64 | | 101.59 | $WF = \sqrt[3.\overline{3}]{\frac{v}{c}}$ | |
| 5 | 125 | | 213.75 | | Speed limit since TNG: "Force of Nature" |

Eq. 6.2

| | | | | | |
|---|---|---|---|---|---|
| 6 | 216 | $WF = \sqrt[3]{\frac{v}{c}}$ Eq. 6.1 | 392.50 | | Normal cruise speed of Federation ships |
| 7 | 343 | | 656.14 | | |
| 8 | 512 | | 1024 | | |
| 9 | 729 | | 1516.38 | | |
| 9.2 | 778.69 | | 1649 | No formula, *Star Trek Encyclopedia* | Max. cruise speed of Galaxy class |
| 9.6 | 884.74 | | 1909 | | Max. safe speed of Galaxy class (96h) |
| 9.9 | 970.30 | | 3053 | | |
| 9.99 | 997.00 | | 7912 | | |
| 9.9999 | 999.97 | | 199516 | | Subspace radio with booster relays |
| 10 | 1000 | | **Infinite** | | Unattainable |
| 11 | 1331 | | Undefined | | |
| 12 | 1728 | | | | |
| 13 | 2197 | | | | |
| 14 | 2744 | | | | |

Tab. 6.1 Comparison of TOS and TNG warp scales

**Background** As the *TNGTM* explains, the real-world reason for changing the warp scale was to put a limit to the warp factors that otherwise may have continued rising indefinitely (in a similar unfortunate fashion as the amounts of quads that can be stored in the computer did on Voyager).

Making up a fictional explanation for the recalibration is not so easy, however. Speed measurements can be expected to be very accurate, their improvement certainly wouldn't lead to a factor of 2 at Warp 9 between the TNG and the TOS scale. Likewise, the peak transitional thresholds that define the integer warp factors wouldn't shift just because the measurement of the ship's power consumption is improved. We have to seek the explanation in a better understanding of subspace physics, one that allowed to refine the working principle of a ship's warp engines in a way that the peak transitional thresholds and hence the warp factors could be moved to higher speeds. In other words, the warp engines were improved in a way that less power was necessary to attain the same speed. This modification must have been so successful and universally applicable that it was decided to change the warp scale. For older ships without this modification the new scale would give us just an equivalent warp factor, meaning that an old ship with the true power expenditure of Warp 6 (TOS scale) is rated with an equivalent warp factor 5 (TNG scale). The same equivalent rating may apply to alien ships whose propulsion system may work differently than that of the Federation, and perhaps not even based on conventional warp fields.

This theory still doesn't explain the dramatic speed increase between Warp 9 and 10 in the TNG scale though. But we could imagine that, as part of the refinement of the propulsion system, at least on Federation vessels, it was recognized that submerging the warp field more than nine layers into subspace was not practically possible or at least inefficient. We have to bear in mind that, as seen in Section 6.2, the maximum possible speed of the TOS era was Warp 8, and that the warp factor 14.1 attained by the Enterprise in TOS: "That Which Survives" may have been just an equivalent figure. So it may have been discovered some time between TOS and TNG that there are other, more efficient methods to attain high speeds than crossing the tenth subspace barrier. And since the peak transitional thresholds 10, 11, etc. don't exist any longer, the scale was changed to reflect this, effectively compressing the whole range from warp factor 9 to infinite warp factor into the new range between Warp 9 and 10. We may even speculate that the refinement of the warp scale has something to do with the Excelsior transwarp experiment, and that "transwarp" here merely means that the principles that would be commonplace in the TNG era were tested for the first time, rather than "transwarp" being something akin to "Borg transwarp" or even Tom Paris' alleged Warp 10 engine.

**Problems** In TNG, DS9 and Voyager there are various statements by crew members such as Data that allow to correlate warp factors, distances and travel times. Sometimes they correspond to Eq. 6.2, sometimes they are significantly off. Overall, however, there are no gross violations of the concept of limited speed, because the ships don't travel to regions of space that should be out of reach, unless the ship's warp drive was explicitly accordingly modified. In Star Trek Voyager it is even a key concept of the series that the ship would need 75 years at maximum warp to travel the 70,000 light-years back to Federation space. This corresponds to about Warp 9.8, going by the figures in Tab. 6.1.

There are three incidents that conflict with the TNG scale more fundamentally because they call the Warp 10 limit into question. The first is in TNG: "Where No One Has Gone Before". During Kosinski's experiment Geordi says at one point: *"We're passing Warp 10."* This should be absolutely impossible considering that Warp 10 means infinite speed, and there can be nothing equal to or even faster than infinite speed. But there is an explanation for Geordi's line. A few moments later Data states that the speed is off the scale. This could mean that Geordi was reading some sort of overflow on the speed display, which was simply leaving the valid range up to Warp 9.9something, so he didn't mean "infinite speed" or even "more than infinite speed" although he effectively said it. Still later, however, Kosinski claims that the "warp barrier" has become meaningless thanks to his work. This seems to corroborate the wrong notion that Warp 10+ might be possible. On the other hand, Kosinski may simply mean the peak transitional thresholds that have shifted or were lowered thanks to his work (or actually, thanks to his assistant, the Traveler).

The second mention of breaking the Warp 10 barrier is in TNG: "Time Squared", where Riker notes that accelerating beyond Warp 10 would enable time travel. More precisely, he states that this constitutes the only known method of going to the past, which is not true for all we know from the countless time travel episodes. Anyway, if we believe him and anything beyond Warp 10 is equivalent to time travel, we may interpret it in a way that achieving exactly Warp 10 may be still impossible. Beyond Warp 10 the speed of a ship becomes undefined within the bounds of our space-time, because speed as a physical quantity always requires a fixed time frame that doesn't exist during a time travel.

The third problem with the Warp 10 limit occurs in the infamous episode VOY: "Threshold". There are many problems and errors with this story, the most pressing of which is how Tom could possibly attain a speed of Warp 10 with his shuttle. As correctly stated in this

very episode by Harry Kim: *"Nothing in the universe can go warp ten. It's a theoretical impossibility. In principle, if you were ever to reach warp ten, you'd be traveling at infinite velocity."* We could argue that, just as in Geordi's case, the speed of the shuttle was simply off scale and hence may have registered as Warp 10, although it was actually slower. The dilemma is that the very intention of Tom's flight was to *"break the warp barrier"* as already Kosinski expressed it, and explicitly to speed up to Warp 10 and not to Warp 9.999something. Also, the speed was later confirmed to have been infinite, unlike it was the case in the TNG episode. This Voyager episode only makes sense if we ignore many statements and essentially major parts of the story.

### 6.4 Possible Future Scale

**Definition** The warp scale changed yet again in the late 24th century, albeit just for the TNG series finale "All Good Things", part of which is set in the year 2395. In this new AGT scale, starships such as the three-nacelled Enterprise-D and the Pasteur routinely achieve Warp 13.

There is no official formula to calculate the speed corresponding to the revised warp factors. In the October 1995 issue of the *Omni* magazine, science advisor Andre Bormanis states: *"I raised that question in a TECH note. Basically, the idea there was that they recalibrated the warp scale. I don't think that ended up in the final draft teleplay, but the idea there was that if you've got ships that can routinely travel at speeds in excess of Warp 9, then maybe it makes sense to recalibrate your speed scale so that Warp 10 is no longer infinite velocity. Maybe Warp 15 will be the ultimate speed limit, and Warp 13 in that scale will be the equivalent of warp 9.95 or something like that."*

See Fig. 6.3 for an illustration of Bormanis' proposal (yellow). Here AGT Warp 13 corresponds to 5200c, which is about the same as TNG Warp 9.95 just as Bormanis reckoned.

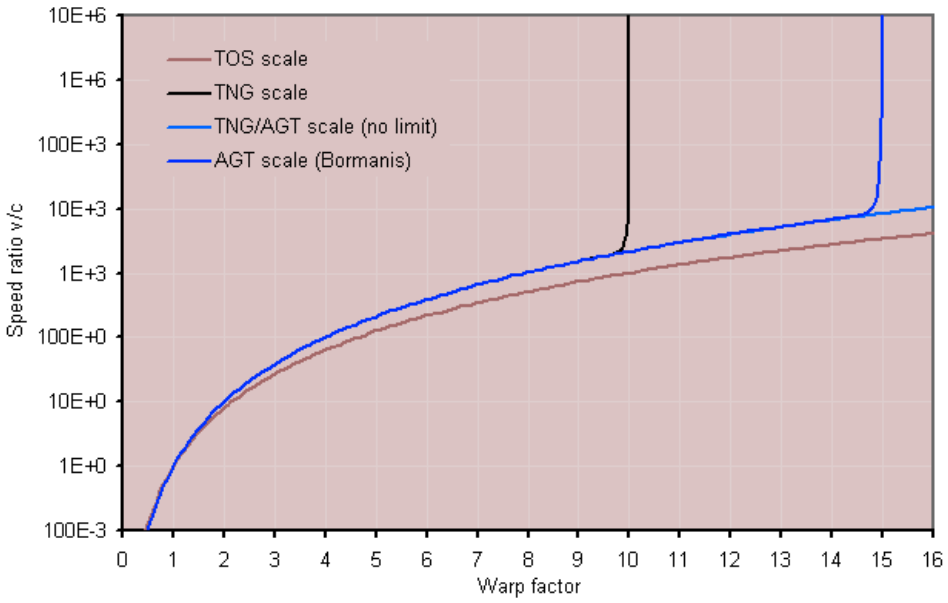



Fig. 6.3 Possible AGT warp scales

**Background** The only thing really known about the AGT scale is that it permits warp factors higher than 10 that, as fans are well aware of, were not allowed in the TNG scale. The obvious real-world reason to switch to this new scale for this one episode was to emphasize that a lot has changed from 2370 to 2395 - just like the additional nacelle on the Enterprise or the new uniforms.

The in-universe explanation may be just as Bormanis suggested. At some point it would be tiresome for the crews of always faster ships to speak out warp factors in the TNG scale with always more decimal places. "Go to Warp 9.9996." and "Go to Warp 9.99996." are unhandy, and they look and sound so much alike to a human crew member that it is likely to become a source of confusion. If, however, the scale was only changed to decompress the warp factor range of high warp speeds for the sake of convenience, the question arises why it was compressed in the first place and if the move of the limit to to Warp 15 or something wouldn't just postpone the problem yet again.

Perhaps, rather than being a totally arbitrary recalibration along the lines "We think Warp 10 is too low a limit, let's move it up to Warp 15", the new AGT scale reflects a change in the propulsion principle yet again. As already explained for the TNG scale in Section 6.3, the highest possible warp factor may denote the deepest possible subspace layer boundary that the ship's warp field penetrates. If, in the year 2395, warp engines are generally modified in a way that they reach down to the the 13th layer, the AGT scale makes sense again. And if 14 subspace layers instead of previously 9 is a commonly accepted limit, the limit of the scale may be located at Warp 15 just as Bormanis suggested, only that the rationale would be different one.

Alternatively, we could imagine that the perhaps impractical warp factor limit was dropped altogether. In Fig. 6.3, the orange curve shows the TNG warp scale without the asymptotic behavior at Warp 10 (TNG scale) or at Warp 15 (AGT scale as suggested by Bormanis).

It is even possible that the warp factors up to Warp 10 are still exactly the same as in the TNG scale, denoting conventional operation of the warp drive, but that a completely new and independent AGT scale begins above Warp 10, denoting a new propulsion system such as transwarp or quantum slipstream drive. In this case, the new warp factors may as well have been named "Transwarp I, II, III, etc." Anyway, the two scales wouldn't be compatible, but we could imagine a transition from one scale to the other. The ship naturally wouldn't go up to Warp 10 with conventional warp drive, but may accelerate up to the break-even point at something like TNG Warp

9.5 and then activate the new propulsion system. After a short transition phase in which both engines (or both working principles of the same engine) are active, the ship may end up at AGT Warp 10, which is perhaps the equivalent to TNG Warp 9.6.

### More about Warp Speed Measurement

- Power & Propulsion - about the right intermix ratio, warp inside a star system, how to stop a starship etc.
- A Close Look at 22nd Century Technology - with considerations about the speed of Enterprise NX-01
- Voyager Inconsistencies - with a rant about Tom's "Warp 10" flight

Warp factor @ Memory Alpha

# 7 Appendix

***The Physics and Technology of Warp Propulsion***

7.1 Glossary - 7.2 Fundamental Constants - 7.3 References - 7.4 Credits

## 7.1 Glossary

*The Physics and Technology of Warp Propulsion*

- Introduction
- Chapter 1 - Real Physics and Interstellar Travel
- Chapter 3 - Subspace
- Chapter 6 - Warp Speed Measurement
- Chapter 7 - Appendix

**Acceleration (a)** Change of velocity v per time t. a=dv/dt

**Accelerator** Machine used to accelerate particles to high speeds (and thus high energy compared to their rest mass-energy).

**Amplitude** In any periodic motion, the maximum displacement from equilibrium.

**Antimatter** Complementary form of matter in which the single particle has the same mass but reversed charge.

**Axiom** Rule without proof, nonetheless valid.

**Bohr, Niels Henrik David** 1885-1962, established a new understanding of the atomic structure, Nobel Prize 1922.

**Center of mass** A point of an object in which its whole mass may be assumed concentrated with respect to outer forces.

**Coherence** Property of a bundle of waves whose relative phases are spatially or temporally equal or constant (example: laser), as opposed to incoherent waves.

**Coordinate system** Diagram which consists of two or more, mostly perpendicular axes (e.g. x and y), in which a function can be illustrated.

**Complex number** Sum of a real part x and an imaginary part x which, since the real axis and the imaginary axis are perpendicular to each other, defines a point (x,y) in the complex plane. s=x+iy

**Constant** Quantity which doesn't change and must not change within an equation.

**Decade** Factor of ten.

**Derivative** Rate of change (steepness) of a function f(x), given as the quotient of two differentials df(x)/dx.

**Differential** Very small (infinitesimal) difference between two values of a quantity.

**Differential equation** Equation which includes one or more derivatives of a variable, and which has to be solved with special methods.

**Diffraction** Bending of a wave passing through a hole, creating a characteristic diffraction pattern in the image plane.

**Dispersion** Variation of the speed of light and therefore the refractive index with wavelength in the same material, resulting in separation of light into its spectrum.

**Doppler shift** Change in wavelength due to relative motion between source and detector.

**Duc de Broglie, Louis Victor Pierre Raymond** 1892-1987, demonstrated that a (small) particle may be represented by a wave, Nobel Prize 1929.

**Efficiency** Mostly the ratio of useful power to total power, also applicable to other quantities.

**Einstein, Albert** 1879-1955, conceived the Special and General Theories of Relativity, Nobel Prize 1921.

**Electromagnetic force** One of the four fundamental forces, arising from to electric charges, both static and moving.

**Electromagnetic wave** Wave consisting of oscillating electric and magnetic fields that move through space at the speed of light.

**Electron ($e^-$)** Elementary particle of small mass and negative charge found in every atom.

**Electron-volt (eV)** Energy gained by an electron which accelerates through a potential difference of one volt, used as an alternative unit of energy.

**Energy (E)** Work stored in matter, for instance as kinetic energy or potential energy.

**Equilibrium** Condition in which the net force on an object is zero.

**Force (F)** Agent that results in accelerating or deforming an object

**Frequency (f)** Number of oscillations per unit time t. f=1/t

**FTL** Acronym for **F**aster **T**han **L**ight.

**Function** Mathematical relationship between two quantities y and x, given as an equation. y=f(x)

**Gravitational field** Distortion of space due to the presence of a mass.

**Gravitational force** One of the four fundamental forces, attraction between two objects due to their mass.

**Heisenberg, Werner** 1901-1976, discovered the Uncertainty Principle, Nobel Prize 1932.

**Hertz (Hz)** Unit of frequency. 1Hz=1/s

**Imaginary number** Multiple of the square root of -1, a number which may be an aid in certain calculations, but which cannot represent a measurable (real) value. i=sqrt(-1)

**Incoherence** Property of a bundle of waves whose relative phases are varying statistically (examples: sunlight, light bulb), as opposed to coherent waves .

**Inertia** Tendency of an object to remain in its current state of motion.

**Infinitesimal** Very small quantity, approaching zero.

**Infrared (IR) radiation** Electromagnetic radiation with a longer wavelength and lower energy content than visible light.

**Integral** Mathematically, the area under a curve f(x), inverse operation to derivation.

**Interference** Superposition of two or more waves, locally producing either larger or smaller amplitudes.

**Ion** Any electrically charged particle, in particular atom nuclei lacking one or more electrons of their nominal complement.

**Ionized** State in which one or more of the usual complement of electrons have left the atom, thereby leaving an ion.

**Joule (J)** Unit of energy (or work or heat). 1J=1Nm

**Kinetic energy** Energy of a mass m due to its motion with a speed v. E_kin=0.5mv^2

**Laser** Acronym for **L**ight **A**mplification by **S**timulated **E**mission of **R**adiation, a light source that produces large amounts of narrow-band light, taking advantage of resonance effects.

**Lorentz, Hendrik Antoon** 1853-1928, developed, among other work, the Lorentz transformation as a basis for Special Relativity, Nobel Prize 1902.

**Mass (m)** Property inherent to any matter, representing its resistance to gravity and acceleration.

**Model** Mathematical description of physical behavior in the form of a set of (simplified) equations.

**Momentum (p)** Product of mass and velocity of an object. p=ma

**Monochromatic light** Light of a single wavelength.

**Neutral** Having a net (electric) charge equal to zero.

**Neutron (n)** Elementary particle with no charge and mass slightly greater than that of a proton.

**Newton (N)** Unit of force. 1N=1kg*m*s^-2

**Newton, Sir Isaac** 1642-1727, conceived, among other important work, the theory of gravitation, thereby explaining the motion of celestial bodies as well as falling objects.

**Nucleon** Proton or a neutron, one of the particles that makes up a nucleus.

**Nucleus** Core of an atom, consisting of protons and neutrons (plural: nuclei).

**Origin** The zero point of a coordinate system.

**Particle** Subatomic object with a definite mass and charge (among other properties).

**Period** Time cycle in which the shape of an oscillation or wave repeats.

**Phase** Fixed shift to a wave, given as an angle

**Photon** Elementary particle which is equivalent to the energy of an electromagnetic wave.

**Planck, Max** 1858-1947, introduced the Quantum Theory, Nobel Prize 1918.

**Planck's constant (h)** Constant determining the relation between the energy E of a photon and its wavelength f. E=hf

**Plasma** Ionized gas.

**Postulate** Assumption necessary to further pursue a theory

**Potential energy** Energy of an object with a mass m due to its position or height h, specifically in a gravitational field with an acceleration g. E_pot=mgh

**Power** Release or consumption of energy E per time t. P=dE/dt

**Principle of superposition** Displacement due to two or more forces, is equal to vector sum of forces.

**Proportional** Changing with the same factor as another quantity.

**Proton (p)** Elementary particle with a positive charge that is nucleus of hydrogen atom.

**Qualitative** Giving a tendency instead of numbers, e.g. "Starship A is faster than starship B".

**Quantitative** Using numbers, e.g. "Starship A travels at 0.38752c".

**Quantum** Smallest discrete amount of any quantity (plural: quanta).

**Quantum mechanics** Study of properties of matter using its wave properties, at very small scales.

**Refraction** Change in direction of light ray when passing from one medium to another.

**Resonance** Effect that occurs when an object is excited with its natural frequency, resulting in a dramatic increase of the amplitude.

**Scalar** Mathematical description of a physical quantity, consisting only of a value, as opposed to a vector.

**Spectrum** Collection of waves with different wavelengths and amplitudes.

**Standing wave** Wave with stationary nodes.

**Statistical** Property of any quantity which is not exactly predictable.

**STL** Acronym for **S**lower **T**han **L**ight.

**Theorem** General scientific rule.

**Thermodynamics** Science of the conversion of one form of energy into another.

**Traveling wave** A moving, periodic disturbance in a medium or field.

**Ultraviolet (UV) radiation** Electromagnetic radiation with a shorter wavelength and higher energy content than visible light.

**Uncertainty principle** Quantum principle that states that it is not possible to know exactly both the position x and the momentum p of an object at the same time.

**Variable** Quantity which is subject to change and supposed to change within an equation.

**Vector** Mathematical description of a physical quantity, consisting of an absolute value (scalar) and a direction.

**Velocity (v)** Change of displacement s per time t. v=ds/dt.

**Visible light** Electromagnetic radiation with a wavelength between 380nm (violet) and 780nm (red).

**Watt (W)** Unit of power. 1W=1J/S

**Wavelength (lambda)** Distance between corresponding points on two successive waves.

**Weight** Force F of gravity g on an object with a mass m. F=mg

**Work (W)** Product of force F and displacement s in the direction of the force. W=Fs

## 7.2 Fundamental Constants

| Name | Symbol | Value | Unit |
|---|---|---|---|
| Speed of light in vacuum | c | 299792458 | $m\ s^{-1}$ |
| Permeability of vacuum | mu_0 | 1.25663706143592e-06 | $N\ A^{-2}$ |
| Permittivity of vacuum | epsilon_0 | 8.854187817e-12 | $F\ m^{-1}$ |
| Newtonian constant of gravitation | G | 6.67259e-11 ± 8.5e-15 | $m^3\ kg^{-1}\ s^{-2}$ |
| Planck constant | h | 6.6260755e-34 ± 4.0e-40 | J s |
| Planck constant in eV | | 4.1356692e-15 ± 1.2e-21 | eV s |
| Planck mass | m_p | 2.17671e-08 ± 1.4e-12 | kg |
| Planck length | l_p | 1.61605e-35 ± 1.0e-39 | m |
| Planck time | t_p | 5.39056e-44 ± 3.4e-48 | s |
| Elementary charge | e | 1.60217733e-19 ± 4.9e-26 | C |
| Electron mass | m_e | 9.1093897e-31 ± 5.4e-37 | kg |
| Electron mass in u | | 0.000548579903 ± 1.3e-11 | u |
| Electron mass in eV | | 510999.06 ± 0.15 | eV |
| Proton mass | m_p | 1.6726231e-27 ± 1.0e-33 | kg |
| Proton mass in u | | 1.00727647 ± 1.2e-08 | u |
| Proton mass in eV | | 938272310 ± 280 | eV |
| Proton-electron mass ratio | | 1836.152701 ± 3.7e-05 | |
| Proton specific charge | | 95788309 ± 29 | $C\ kg^{-1}$ |
| Neutron mass | m_n | 1.6749286e-27 ± 1.0e-33 | kg |
| Neutron mass in u | | 1.008664904 ± 1.4e-08 | u |
| Neutron mass in eV | | 939565630 ± 280 | eV |
| Neutron-electron mass ratio | | 1838.683662 ± 4.0e-05 | |
| Neutron-proton mass ratio | | 1.001378404 ± 9e-09 | |
| Avogadro constant | N_A | 6.0221367e+23 ± 3.6e+17 | $mol^{-1}$ |
| Boltzmann constant | k | 1.380658e-23 ± 1.2e-28 | $J\ K^{-1}$ |
| Boltzmann constant in eV | | 8.617385e-05 ± 7.3e-10 | $eV\ K^{-1}$ |
| Molar volume ideal gas, STP | V_m | 0.0224141 ± 1.9e-07 | $m^3\ mol^{-1}$ |
| Electron volt | eV | 1.60217733e-19 ± 4.9e-26 | J |
| Atomic mass unit | u | 1.6605402e-27 ± 1.0e-33 | kg |
| Standard acceleration of gravity | g_n | 9.80665 | $m\ s^{-2}$ |

Tab. 7.1 Fundamental physical constants

This data is derived from the 1986 CODATA recommended values of the fundamental physical constants. Complete list of constants: NIST Reference.

## 7.3 References

[Use] Usenet Relativity FAQ, Desy mirror

[Ast] Astronomy FAQ, http://www.faqs.org/faqs/astronomy/faq/

[Sce] B. Scent, S. Davis, L. Scheinbeim, Black Holes - Portals into the Unknown

[Mil] M. G. Millis, Warp Drive When?

[Hin1] J. Hinson, Relativity and FTL Travel

[Hin2] J. Hinson, Subspace Physics

[Bel] J. Bell, Star Trek Technology: Frequently Asked Questions Lists

[STTR] R. Mercer, The Star Trek Technical Reference

[DITL] G. Kennedy, Daystrom Institute Technical Library

[STD] C. Rühl, Star Trek Dimension - Subspace Manual

[Alc94] M. Alcubierre, *The warp drive: hyper-fast travel within general relativity*, Classical and Quantum Gravity, Vol. 11, L73-77, May

1994, link to PDF

[Ein92] A. Einstein, *Über die spezielle und die allgemeine Relativitätstheorie*, reprint, Vieweg, 1992

[Fey63] R. P. Feynman, R. B. Leighton, M. Sands, *The Feynman Lectures on Physics*, Vol. I, Addison-Wesley, 1963

[Ger89] C. Gerthsen, H. Kneser, H. Vogel, *Physik*, 16. Auflage, Springer Verlag, 1989

[Haw98] S. W. Hawking, *Eine kurze Geschichte der Zeit*, reprint, Rowohlt, 1998

[Kra96] L. M. Krauss, *Die Physik von Star Trek*, Heyne Verlag, 1996

[Oku99] M. Okuda, D. Okuda, D. Drexler, *The Star Trek Encyclopedia*, Pocket Books, 1999

[Sex87] R. and H. Sexl, *Weiße Zwerge - Schwarze Löcher*, 2. Auflage, Vieweg, 1987

[Ste91] R. Sternbach, M. Okuda, *Star Trek: The Next Generation Technical Manual*, Pocket Books, 1991

[Whi68] Stephen E. Whitfield, Gene Roddenberry, *The Making of Star Trek*, Del Rey Books, 1968

[Zim98] H. Zimmerman, R. Sternbach, D. Drexler, I. S. Behr, *Star Trek: Deep Space Nine Technical Manual*, Pocket Books, 1998

### 7.4 Credits

Thanks to Jörg, to Tobias Heintz and to Jan for suggestions and corrections.